IMPROVING AGE-BASED FISH STOCK ASSESSMENTS, WITH AN APPLICATION TO AMERICAN PLAICE ON THE GRAND BANK OF NEWFOUNDLAND

by © Andrea M.J. Perreault

A Thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Master of Science Department of Fisheries Science

Memorial University of Newfoundland

May 2020

St. John's, Newfoundland and Labrador

MEMORIAL UNIVERSITY

DEPARTMENT OF FISHERIES SCIENCE

The undersigned hereby certify that they have read and recommend to the Faculty of Fisheries Science for acceptance a thesis entitled "Improving age-based fish stock assessments, with an application to American plaice on the Grand Bank of Newfoundland" by Andrea Perreault in partial fulfillment of the requirements for the degree of Master of Applied Science.

Dated: May 16th

Supervisor: _____
Dr. Noel Cadigan

Reader: _____

MEMORIAL UNIVERSITY

Date: May 16th

Author: Andrea Perreault

Title: Improving age-based fish stock assessments, with an application to American plaice on the Grand Bank of Newfoundland

Department or School: Fisheries Science

Degree: M.A.Sc. Convocation: May Year: 2020

Permission is herewith granted to Memorial University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.

_____

Signature of Author

Table of Contents

Abstract **vi**

Acknowledgements **vii**

List of Tables **vii**

List of Figures **x**

Chapter 1 Introduction **1**

1.1 Age-based stock assessment models 1

1.2 Data and gaps 21

1.3 Overview and Objectives 24

Chapter 2 Estimation of growth parameters based on length-stratified age samples **29**

2.1 Introduction 29

2.2 Materials and Methods 33

2.3 Results 48

2.4 Discussion 54

2.5 Acknowledgments 60

Bibliography **62**

Chapter 3 A state-space stock assessment model for American plaice on the Grand Bank of Newfoundland **66**

3.1 Introduction 66

3.2 Materials and Methods 68
3.3 Results 77
3.4 Discussion 84
3.5 Acknowledgements 87

Bibliography **88**

Chapter 4 Summary and Papers **91**

4.1 Paper 1 92
4.2 Paper 2 105

Bibliography **128**

Appendix A Derivation of density function for second phase length-at-age data **136**

Appendix B Horvitz-Thompson weights **137**

Appendix C Neglecting spatial structure of sampling scheme **138**

Appendix D Supporting figures for state-space stock assessment model for 3LNO American plaice **139**

Appendix E Data inputs for state-space stock assessment model for 3LNO American plaice **147**

Appendix F Exploratory model (EM) fitting **154**

Appendix G SSM supporting code **193**

Abstract

Age-based fish stock assessment models use available data to provide fisheries managers with estimates of population processes and to determine sustainable harvest rates. However, the data to inform these models is often complex and ignoring or oversimplifying these complexities can result in unsustainable or sub-optimal harvesting advice. This thesis improves age-based stock assessment models by accounting for important variability in the data. I first simulation tested nine methods that aim to account for the commonly used length-stratified age sampling design when estimating growth parameters. My results showed that commonly used methods had poor accuracy and the empirical proportion approach was optimal. Secondly, I developed a state-space stock assessment model for American plaice that allowed for errors in the underlying population processes and provided improvements in the retrospective plots. This research improved our understanding of American plaice population dynamics and growth and is a step forward in fitting more realistic integrated age-based stock assessment models.

Acknowledgements

I wish to express my sincere appreciation to my supervisor Dr. Noel Cadigan for his guidance and expertise throughout the entirety of this research. As his student, I have become a better writer, researcher and overall scientist and I am grateful to have had his support and direction throughout this project. I also thankfully acknowledge the funding for this research that was provided by the Ocean Frontier Institute, through an award from the Canada First Research Excellence Fund, the Ocean Choice International Industry Research Chair program at the Marine Institute of Memorial University of Newfoundland and from a Natural Sciences and Engineering Research Council of Canada Master's Graduate Scholarship. Many thanks are also extended to Dr. Anders Nielsen, Danish Technical University, for advice on more computationally efficient ways to implement state-space stock assessment models in TMB.

A huge thanks and appreciation to my husband Curtis, Abs, and my sisters Genn, Meara, Amissa, Jess and Chelsea who have been my emotional support and cheerleaders throughout this process. I owe the majority of my sanity and perseverance throughout this work to having had such an incredible support system.

List of Tables

2.1 Log-likelihood functions used to estimate vonB population parameters, dbin represents the binomial probability mass function, pbin the cumulative binomial probability mass function, crl is the continuation ratio logit transformation, and equil. is the equilbrium approximation in equation 2.8. 36

2.2 Distributions and models used to generate population simulations; min = 0.10 and max = 0.50 for the uniform distribution; truncNorm is the truncated normal distribution with min = 0, max = ∞, to prevent generating negative values; when a_0 was estimated freely, a_0 was used, otherwise L_0 was used. 45

2.3 RRMSEs (relative root mean squared error) and RelBias (relative bias) for vonB parameter estimates from 1000 simulations with length at age 0 fixed at 3 cm, for various length bin sizes. The abbreviations for the methods are: PF (partial-full), BC (bias-corrected mean length-at age), WBC (weighted BC), HT (Horvitz-Thompson weighted), CW (calibrated weighted) and EP (empirical proportion). 49

2.4 RRMSEs (relative root mean squared error) and RelBias (relative bias) for vonB parameter estimates from 1000 simulations with a_0 estimated freely, for various length bin sizes. The abbreviations for the methods are: PF (partial-full), BC (bias-corrected mean length-at age), WBC (weighted BC), HT (Horvitz-Thompson weighted), CW (calibrated weighted) and EP (empirical proportion). 50

2.5 RRMSEs (relative root mean squared error) and RelBias (relative bias) for vonB parameter estimates from 1000 simulations with a_0 estimated freely for 1cm bin sizes, target stratum sample size $m_k = 50$, and simulated population size (N) or five times simulated population size ($5N$). The abbreviations for the methods are: PF (partial-full), BC (bias-corrected mean length-at age), WBC (weighted BC), HT (Horvitz-Thompson weighted), CW (calibrated weighted) and EP (empirical proportion). . . . 51

2.6 Parameter estimates and standard errors (SEs) for vonB model parameters for female American plaice data in NAFO Divisions 3LNO in 2014. The abbreviations for the methods are BC (bias-corrected mean length-at age), WBC (weighted BC) and EP (empirical proportion) 52

3.1 Distributions and models used to generate population simulations; min = 0.10 and max = 0.50 for the uniform distribution; truncNorm is the truncated normal distribution with min = 0, max = ∞, to prevent generating negative values; when a_0 was estimated freely, a_0 was used, otherwise L_0 was used. 73

3.2 State-space model fixed effect parameters 75

3.3 Model estimates of variance parameters and some population parameters with percent coefficient of variation. (M1 is base model, M2 is base model with natural mortality=0.20, M3 is base model with natural mortality increased by 0.20, M4 is base model with upper catch bounds set to half the base model upper bounds, M5 is base model with catchability freely estimated). . 80

List of Figures

2.1 Diagram of steps used in population simulations; LSAS is length-stratified age sampling. 44

2.2 Estimated vonB growth model for female American plaice data in Divisions 3LNO in 2014. The points have added transparency to better represent the density of the observations. The various fits are: blue line (random), green line (EP), yellow line (Candy), purple line (bias-corrected mean length-at age (BC)) and orange line (WBC). 54

2.3 Plot of simulated population N with estimated vonB growth model from sub-sample n. The points have added transparency to better represent the density of the observations. The various fits are: blue line (random), green line (EP), yellow line (Candy), purple line (bias-corrected mean length-at age (BC)), orange line (WBC) and red line (population growth model) . . 55

2.4 Plot of simulated sub-population n with estimated vonB growth model from sub-sample. The points have added transparency to better represent the density of the observations. The various fits are: blue line (random), green line (EP), yellow line (Candy), purple line (bias-corrected mean length-at age (BC)), orange line (WBC) and red line (population growth model). . 56

3.1 Model estimated population abundance, spawning stock biomass, average fishing mortality rates (ages 9-14) for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent the 95% confidence intervals. 78
3.2 Model estimated log catch numbers for ages 5-15+ (solid line) for American plaice in NAFO Divisions 3LNO, the shaded grey represents the region between the log lower catch bounds and the log upper catch bounds. 78
3.3 Survey catchability patterns for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. A separate catchability parameter is estimated for two gear types (Engel and Campelen) for the spring and fall surveys for ages 1-4. 79
3.4 Model predicted process errors at age for American plaice in NAFO Divisions 3LNO. 81
3.5 Retrospective estimates from 2008-2017 for total abundance, spawning stock biomass, average fishing mortality rates (ages 9-14) and recruitment for years 1960-2017 for American plaice in NAFO Divisions 3LNO. 82
3.6 Model (SSM) and VPA estimated spawning stock biomass for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent the 95% confidence intervals. 82

3.7 Model (SSM) and VPA estimated average fishing mortality rates (ages 9-14) for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent the 95% confidence intervals. 83

3.8 Model (SSM) and VPA retrospective estimates from 2013-2017 of average fishing mortality rates (ages 9-14) and of spawning stock biomass for years 2014-2017 for American plaice in NAFO Divisions 3LNO. 83

3.9 Relative difference from model and self-simulated sample for spawning stock biomass and average fishing mortality rates (ages 9-14). The solid grey line in the median of the estimates and shaded grey regions represent the lower 10% and upper 90% bounds. 84

D.1 Model predicted survey fits (red) and observed survey indices at age (black) for ages 1-7 for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. 139

D.2 Model predicted survey fits (red) and observed survey indices at age (black) for ages 8+ for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. 140

D.3 Standardized bubble plot of residuals at age for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative. 141

D.4 Standardized residuals for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. 142

D.5 Model predicted catch proportions at age (red) vs observed catch proportion at age (black) for American plaice in NAFO Divisions 3LNO. 143
D.6 Standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO. 144
D.7 Bubble plot of standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO. 145
D.8 Model predicted log fishing mortality rates for American plaice in NAFO Divisions 3LNO. 145
D.9 Model predicted log fishing mortality deviations for American plaice in NAFO Divisions 3LNO. 146
D.10 Model predicted fishing mortality rates for American plaice in NAFO Divisions 3LNO. 146

E.1 Abundance data for ages 1-7 for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. . . 147
E.2 Abundance data for ages 8+ for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. . . 148
E.3 Standardized proportion by age across years for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative and black is zeros 149
E.4 Mean standardized survey indices for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. Red is fall, green is Spanish and blue is spring survey. 150

E.5 Landings estimates for American plaice in NAFO Divisions 3LNO. 151
E.6 Catch proportions at age for American plaice in NAFO Divisions 3LNO. 151
E.7 Proportion mature at age for American plaice in NAFO Divisions 3LNO. 152
E.8 Stock weights at age for American plaice in NAFO Divisions 3LNO. 152
E.9 Catch weights at age for American plaice in NAFO Divisions 3LNO. 153

F.1 EM1 estimated population abundance, spawning stock biomass, average fishing mortality rates (ages 9-14) and recruitment for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent the 95% confidence intervals. 155
F.2 EM1 predicted survey fits (red) and observed survey indices at age (black) for ages 1-7 for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. 156
F.3 EM1 predicted survey fits (red) and observed survey indices at age (black) for ages 8+ for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. 157
F.4 EM1 standardized bubble plot of residuals at age for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative. 158

F.5 EM1 standardized residuals for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. . . 159
F.6 EM1 estimated log catch numbers for ages 5-15+ (solid line) for American plaice in NAFO Divisions 3LNO, the shaded grey represents the region between the log lower catch bounds and the log upper catch bounds. 160
F.7 EM1 predicted catch proportions at age (red) vs observed catch proportion at age (black) for American plaice in NAFO Divisions 3LNO. 161
F.8 EM1 standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO. 162
F.9 EM1 bubble plot of standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO. . . . 163
F.10 EM1 predicted log fishing mortality rates for American plaice in NAFO Divisions 3LNO. 163
F.11 EM1 predicted log fishing mortality deviations for American plaice in NAFO Divisions 3LNO. 164
F.12 EM1 predicted fishing mortality rates for American plaice in NAFO Divisions 3LNO. 165
F.13 EM1 predicted process errors at age for American plaice in NAFO Divisions 3LNO. 166
F.14 EM1 bubble plot of predicted process errors at age for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative. 166

F.15 EM1 survey catchability patterns for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. A separate catchability parameter is estimated for two gear types (Engel and Campelen) for the spring and fall surveys for ages 1-4. 167
F.16 EM1 (SSM) and VPA estimated spawning stock biomass for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent the 95% confidence intervals. 168
F.17 EM1 (SSM) and VPA estimated average fishing mortality rates (ages 9-14) for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent the 95% confidence intervals. 168
F.18 EM2 estimated population abundance, spawning stock biomass (SSB), average fishing mortality rate (ages 9-14) and recruitment for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent the 95% confidence intervals. 169
F.19 EM2 predicted survey fits (red) and observed survey indices at age (black) for ages 1-7 for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. 170
F.20 EM2 predicted survey fits (red) and observed survey indices at age (black) for ages 8+ for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. 171
F.21 EM2 standardized bubble plot of residuals at age for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative. 172

F.22 EM2 standardized residuals for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. . . 173
F.23 EM2 estimated log catch numbers for ages 5-15+ (solid line) for American plaice in NAFO Divisions 3LNO, the shaded grey represents the region between the log lower catch bounds and the log upper catch bounds. 173
F.24 EM2 predicted catch proportions at age (red) vs observed catch proportion at age (black) for American plaice in NAFO Divisions 3LNO. 174
F.25 EM2 standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO. 175
F.26 EM2 bubble plot of standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO. . . . 176
F.27 EM2 predicted log fishing mortality rates for American plaice in NAFO Divisions 3LNO. 176
F.28 EM2 predicted log fishing mortality deviations for American plaice in NAFO Divisions 3LNO. 177
F.29 EM2 predicted fishing mortality rates for American plaice in NAFO Divisions 3LNO. 177
F.30 EM2 predicted process errors at age for American plaice in NAFO Divisions 3LNO. 178
F.31 EM2 bubble plot of predicted process errors at age for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative. 178

F.32 EM2 survey catchability patterns for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. A separate catchability parameter is estimated for two gear types (Engel and Campelen) for the spring and fall surveys for ages 1-4. 179
F.33 EM2 (SSM) and VPA estimated spawning stock biomass for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent the 95% confidence intervals. 179
F.34 EM2 (SSM) and VPA estimated average fishing mortality rates (ages 9-14) for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent the 95% confidence intervals. 180
F.35 EM3 estimated population abundance, spawning stock biomass (SSB), average fishing mortality rate (ages 9-14) and recruitment for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent the 95% confidence intervals. 181
F.36 EM3 predicted survey fits (red) and observed survey indices at age (black) for ages 1-7 for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. 182
F.37 EM3 predicted survey fits (red) and observed survey indices at age (black) for ages 8+ for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. 183
F.38 EM3 standardized bubble plot of residuals at age for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative. 184

F.39 EM3 standardized residuals for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. . . 185
F.40 EM3 estimated log catch numbers for ages 5-15+ (solid line) for American plaice in NAFO Divisions 3LNO, the shaded grey represents the region between the log lower catch bounds and the log upper catch bounds. 185
F.41 EM3 predicted catch proportions at age (red) vs observed catch proportion at age (black) for American plaice in NAFO Divisions 3LNO. 186
F.42 EM3 standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO. 187
F.43 EM3 bubble plot of standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO. . . . 188
F.44 EM3 predicted log fishing mortality rates for American plaice in NAFO Divisions 3LNO. 188
F.45 EM3 predicted log fishing mortality deviations for American plaice in NAFO Divisions 3LNO. 189
F.46 EM3 predicted fishing mortality rates for American plaice in NAFO Divisions 3LNO. 189
F.47 EM3 predicted process errors at age for American plaice in NAFO Divisions 3LNO. 190
F.48 EM3 bubble plot of predicted process errors at age for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative. 190

F.49 EM3 survey catchability patterns for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. A separate catchability parameter is estimated for two gear types (Engel and Campelen) for the spring and fall surveys for ages 1-4. 191
F.50 EM3 (SSM) and VPA estimated spawning stock biomass for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent the 95% confidence intervals. 191
F.51 EM3 (SSM) and VPA estimated average fishing mortality rates (ages 9-14) for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent the 95% confidence intervals. 192

Chapter 1

Introduction

1.1 Age-based stock assessment models

Commercial fishing is an essential Canadian economic activity and is most often managed via estimates of population processes (e.g. abundance, fishing mortality rates, biomass) that are predicted from fish stock assessment models. These statistical models use available data to produce estimates of stock size and harvest rates, and are also used to determine sustainable harvest rates and lower bounds on stock size (i.e. limit reference points) that management actions should avoid with high probability. Stock assessment models range in complexity based on the availability of data, from simple catch-only models that rely on commercial landings data (e.g. MacCall 2009), to integrated state-space models that incorporate survey, catch-at-age composition, tagging and landings data (e.g. Cadigan 2015). When age composition data are available, age-based methods are preferred, as the information provided by the growth, size and structure of the stock can greatly influence how the fishery is managed (Haddon 2010).

Two fundamental equations serve as the foundation for age-based fish stock assessment models: the cohort equation and the catch equation. First, let a denote age in the model, where the range of model ages is $a = 1, ...A$ and A is the maximum model age. Similarly, let y denote a year, and the model years range from $y = 1, ..., Y$. The cohort equation is

$$N_{a,y} = N_{a-1,y-1} \exp^{-Z_{a-1,y-1}}, \tag{1.1}$$

where $N_{a,y}$ is the cohort abundance at age a and year y, $Z_{a,y} = M_{a,y} + F_{a,y}$ is the total mortality rate, described by the sum of the natural mortality rate $M_{a,y}$ (i.e. all mortality other than fishing mortality) and the fishing mortality rate $F_{a,y}$[1]. The numbers caught by the fishery at each age and year (i.e. $C_{a,y}$) are given by the catch equation, (Baranov 1918),

$$C_{a,y} = \frac{F_{a,y}}{Z_{a,y}} N_{a,y}(1 - \exp^{-Z_{a,y}}). \tag{1.2}$$

A basic requirement for all models described below is commercial catch-at-age data, which are not directly observed but based on landings estimates from the various fishing fleets and sub-sampling of their catches for age and length measurements (see, e.g., Hilborn & Walters 2013). Catches are sub-sampled for ageing because in practice it is often too expensive or unrealistic to age all fish in the catch. Data-related issues, including ignoring the variability in the catch-at-age estimates, are discussed in Section 1.4.

Early research found that catch-at-age data alone are often not enough to effectively estimate model parameters (e.g. Doubleday 1976, Pope & Shepherd 1982) and auxiliary information (e.g. survey abundance indices) needs to be included in the model formulation to resolve this issue. Auxiliary data are most often derived from two sources: fishery independent and fishery dependent data. Fishery dependent data are typically collected from commercial landings, fisheries logbooks and on-board commercial observers and are used to derive estimates of fishing effort, abundance indices and biological characteristics of the population (e.g. length, stomach contents). Fishery independent data are typically collected via standardized methods (e.g. stratified sampling) through governmental and/or scientific programs,

[1]If it is difficult to age fish at older ages or it is unreasonable to assume that all fish die out at the oldest age, a plus group is often used to model the fish of the oldest ages, i.e., $N_{y,A^+} = N_{y-1,A^+-1} \exp^{-Z_{y-1,A^+-1}} + N_{y-1,A^+} \exp^{-Z_{y-1,A^+}}$, where A^+ represents the oldest ages grouped together. For simplicity, I present model formulations that do not make the assumption of a plus group.

most often through research survey sampling and tagging programs. Fishery independent data are often used to derive indicators of population abundance and biological characteristics of the population as they are considered more informative (i.e. with less bias and uncertainty) than fishery dependent data.

In the following sections, I describe the evolution of age-based stock assessment models, beginning with the simple models that rely only on commercial catch-at-age data, to more complex integrated models that also incorporate various sources of auxiliary data.

Virtual population analysis

Historically, the most commonly implemented age-based stock assessment model was virtual population analysis (VPA), which was popularized by Gulland (1965) and based on the earlier work of Fry (1949). VPA follows from the logic that if the catch-at-age data are assumed to be known without error in year y, then we have a minimum estimate of how many fish were alive in year $y-1$; that is, there has to be at least as many in the stock as were caught. If $M_{a,y}$ is also assumed known, we can back-calculate from the cohort equation (1.1) to obtain an abundance estimate at the youngest age. This backwards method requires initial estimates of abundance at the oldest age, derived from the rearranged catch equation (1.2),

$$N_{A,y} = \frac{C_{A,y}}{\frac{F_{A,y}}{Z_{A,y}}(1-\exp^{-Z_{A,y}})}. \tag{1.3}$$

Here, (1.3) requires an independent estimate of $F_{A,y}$, called the terminal F. $C_{a,y}$ represents fish of age a caught up until the end of time period y, and $N_{a+1,y+1}$ represents the stock size available to the fishery in the beginning of year $y+1$. Thus both $C_{a,y}$ and $N_{a+1,y+1}$ reference the same time period and can be combined to obtain

the ratio of the population abundance to the catch (i.e., by combining 1.1 and 1.2),

$$\frac{N_{a+1,y+1}}{C_{a,y}} = \frac{Z_{a,y} \exp^{-Z_{a,y}}}{F_{a,y}(1 - \exp^{-Z_{a,y}})}. \tag{1.4}$$

For $a = A - 1$, $N_{A,y}, C_{A-1,y-1}$ and $M_{A-1,y-1}$ are known, and (1.4) can be solved for $F_{A-1,y-1}$. Then, $F_{A-1,y-1}$ and $N_{A,y}$ can be substituted into the cohort equation (1.1) to estimate $N_{A-1,y-1}$. This process is repeated until the youngest age is reached.

Equation (1.4) is nonlinear in $F_{a,y}$, thus an iterative procedure is required to solve the related equations. At the time of Gulland (1965)'s publication this was very computationally intensive and VPA was rarely implemented in practice. Pope (1972) proposed a simplification, often called cohort analysis, that greatly increased the applicability of VPA. This method assumes that the entire catch is taken exactly midway through the year via the discrete approximation,

$$\exp^{M_{a,y}/2} = \frac{Z_{a,y}(1 - \exp^{-F_{a,y}})}{F_{a,y}(1 - \exp^{-Z_{a,y}})},$$

where $M_{a,y}$ is assumed constant and known across all ages and years. As in Gulland's VPA, Pope's method calculates terminal abundance from (1.3) and still requires an external estimate of the terminal F (see Xiao & Wang (2007) for a review and derivations of Pope's equations), but with the cohort equation now given by

$$N_{a,y} = N_{a+1,y+1} \exp^{M_{a,y}} + C_{a,y} \exp^{-M_{a,y}/2},$$

and $F_{a,y}$ given by

$$F_{a,y} = \log\Big(\frac{N_{a,y}}{N_{a+1,y+1}} - M_{a,y}\Big).$$

Thus, the abundance estimates and fishing mortality rates can be calculated directly without the need for an iterative procedure.

Murphy (1965) developed a similar method to Gulland's VPA using a "forward" solution that starts from the abundance at the youngest ages and moves forward in time. Early research comparing the two methods found that estimates of fishing mortality rates from the backwards method converged to their true values, whereas those from the forwards method did not (Jones 1961, Tomlinson 1970). For brevity, I omit the details and refer the reader to Megrey (1989), who gives an excellent review of age-based stock assessment models, including Murphy's forward method.

The parameters to estimate in the VPA model are the numbers at the oldest age $N_{A,y}$ (or equivalently $F_{A,y}$) and the numbers in the last year $N_{a,Y}$. The number of parameters to estimate are identical or greater than the number of data points and a major concern surrounding these methods is that there are infinitely many "solutions" that depend entirely on the choice of terminal F (see, e.g. Megrey 1989). To reduce the number of parameters to estimate F-constraints on the oldest age (e.g. $F_{A,y} = \text{ave}(F_{a1,y} - F_{a2,y})$) can be used to provide values for the $N_{A,y}$'s (see, e.g. Gavaris 1988). For large F's, the $N_{A,y}$'s are typically very small and the models are not sensitive to the F-constraints used. This is not the case for the $N_{a,Y}$'s since these numbers (i.e. the survivors) are not small. However, F constraints have been used for some of the ages in the last year to further reduce the number of unknown parameters (e.g. if $A = 15$, then $F_{15,y} = \text{ave}(F_{10,y} - F_{12,y})$ and in the last year $F_{14,Y} = \text{ave}(F_{9,Y} - F_{11,Y})$ and $F_{13_Y} = \text{ave}(F_{8,Y} - F_{10,Y})$) with partial recruitment (i.e. the age pattern in fishing mortality rates) used as guidance on reasonable F-constraints (see, e.g. R. Rideout & Brattey 2017). However, the method of F-constraints cannot be used when there are zeros in the catch.

Another approach to reducing the number of parameters to estimate in the VPA method is via the separability assumption. This method is somewhat linked to looking at the fish partial recruitment, since the partial recruitment (i.e. $F_{a,y}/\max_a(F_{a,y})$)

is expected to change smoothly over ages and years. The separability assumption is often attributed to Doubleday (1976), however Megrey (1989) credits Agger et al. (1971)'s earlier paper with the first recorded implementation. The separability assumption treats the fishing mortality rate as the product of an age-dependent fisheries selectivity s_a (i.e. the how susceptible a fish of age a is to the fishing gear) and an age-independent fishing intensity f_y,

$$F_{a,y} = f_y s_a, \tag{1.5}$$

which effectively reduces the number of parameters to be estimated. To ensure that fishing mortality rates are well-defined, a normalization constant is required. Most commonly, $s_a = 1$ for some age that is assumed to be fully exploited by the fishery.

Pope & Shepherd (1982) were the first to incorporate the separability assumption in VPA by combining (1.5) and the log of the catch equation (1.2),

$$\log\left[\frac{C_{a+1,y+1}}{C_{a,y}}\right] = \log\left[\frac{\exp^{-Z_{a,y}} f_{y+1} s_{a+1} Z_{a,y}(1-\exp^{-Z_{a+1,y+1}})}{f_y s_a Z_{a+1,y+1}(1-\exp^{-Z_{a,y}})}\right] + \epsilon_{CR},$$

where $\epsilon_{CR} \sim N(0, 2\sigma^2_{CR})$. Parameters are estimated using a sequential two stage least squares algorithm, which I do not detail here, but refer to Megrey (1989) for a detailed explanation. Research soon showed that catch-at-age data were not enough to reliably estimate stock abundance as estimates of fishing mortality rates and stock abundance are highly negatively correlated (Deriso et al. 1985). Incorporating auxiliary information to help identify fishing mortality rates was the next step forward. Many "ad hoc" methods were proposed in the early implementation of auxiliary information and Pope & Shepherd (1985) provide a thorough study of these methods. Two VPA software packages that incorporate auxiliary information and are still in use today are described briefly below.

Gavaris (1988) popularized the adaptive-framework (ADAPT), which supplements the catch-at-age data with at least one stock size (e.g. biomass, catch-per-unit-effort) or fishing mortality rate index. Most often, standardized catch-per-unit effort data from surveys and/or commercial data are used. Estimates of the terminal abundance and fishing mortality at each age and year are found by minimizing the sum of the squared differences of the predicted and observed auxiliary index. The ADAPT software has been available since the late 1980's (Lassen & Medley 2001) and has enabled the straight-forward implementation of the ADAPT-VPA method for many stocks worldwide, including American plaice (*Hippoglossoides platessoides*) on the Grand Bank of Newfoundland (Wheeland et al. 2018) and Georges Bank and North Atlantic haddock (*Melanogrammus aeglefinus*, Brooks 2017, DFO 2017) today.

Extended survivor analysis (XSA) is an approach that was developed to avoid the issue of "ad hoc" tuning and to allow for errors in the terminal abundance (Cotter et al. 2004). XSA treats the catch-at-age data as exact and can only allow for age dis-aggregated auxiliary indices (Darby & Flatman 1994, Shepherd 1999). The model is iteratively tuned to the auxiliary data and does not require optimization to solve the system of equations. The XSA software package was once one of the most commonly used stock assessment tools in Europe (ICES 2010), but more realistic model formulations (e.g. SAM, detailed later) are now being implemented in its place. However, XSA is still the main assessment tool for many stocks today, including sole (*Solea solea*) on the Eastern English Channel (ICES 2019b) and western English Channel/southern Celtic Seas cod (*Gadus morhua*, ICES 2019a).

Statistical catch-at-age

Doubleday (1976) has been credited with formulating the first stochastic age-based stock assessment model, which predicts numbers at age forward in time starting at the

youngest age. Let r be the age that a fish is available to the fishery (i.e. recruitment), then the cohort equation (1.1) is expressed in terms of fish born in the same year, called the year-class (i.e. $y - r$). For age a fish in year y, the cohort model is

$$N_{a,y} = N_{r,(y-a+r)} \exp^{-\sum_{j=1}^{a-r} Z_{a-j,y-j}}. \tag{1.6}$$

Thus, the abundance at age a in year y can be derived from the recruitment abundance and subsequent mortality rates. Doubleday combined the year-class equation (1.6), the catch equation (1.2), and the separability assumption (1.5),

$$\log(C_{a,y}) = \log\left[\frac{f_y s_a}{f_y s_a + M_{a,y}}(1-\exp^{-f_y s_a + M_{a,y}}) N_{r,y-a+r} \exp^{-\sum_{j=1}^{a-r} f_{y-j} s_{a-j} + M_{a-j,y-j}}\right] + \epsilon_{a,y},$$

where ϵ is assumed to be independent and normally distributed with mean zero and constant variance. As previously defined, $Z_{ay} = f_y s_a + M_{ay}$, where f_y and s_a are parameters to estimate and M_{ay} is assumed to be known. To estimate parameters, this method minimizes the sum of squares between the observed catch-at-age $\log(C_{a,y})$, and the predicted catch at age $\log(\hat{C}_{a,y})$, i.e.,

$$SSR_C = \sum_{y=1}^{Y} \sum_{a=1}^{A} \left[\log(\hat{C}_{a,y}) - \log(C_{a,y})\right]^2.$$

Doubleday applied an iterative linear approximation and estimation to obtain the least squares estimates and required a second equation to generate realistic starting values (see Megrey 1989, for details).

The catch-at-age data do not contain enough information to precisely estimate fishing mortality rates (Doubleday 1976, Pope & Shepherd 1982), and Paloheimo (1980) developed a solution to this problem by incorporating auxiliary information via fishery-dependent data. Paloheimo (1980) made the simplifying assumption that

fishing mortality is proportional to yearly fishing effort, i.e.,

$$F_y = qE_y, \tag{1.7}$$

where q is the catchability coefficient, assumed constant across years and E_y is the observed or measured fishing effort. Paloheimo also greatly simplified calculations and increased the ease of applicability of his method by proposing the approximation,

$$\log\Big(\frac{1-\exp^{-Z_{a,y}}}{Z_{a,y}}\Big) \approx -\frac{Z_{a,y}}{2}. \tag{1.8}$$

The model then combines the catch equation (1.2), the year-class equation (1.6), fishing effort (1.7), and the simplification (1.8),

$$\log\Big(\frac{C_{a,y}}{E_y}\Big) = \log\Big[-\frac{M_{a,y}+qs_aE_y}{2} qN_{r,y-a+r}\exp^{-\sum_{j=1}^{a-r}-(M_{a,y}+qs_{a-j}E_{y-j})}\Big] + \epsilon_{a,y}, \tag{1.9}$$

where ϵ are assumed independent and normally distributed with mean zero and constant variance. If all ages are assumed to be fully recruited, i.e. $s_a = 1$ for all ages, then (1.9) reduces to a simple linear equation and the parameters are estimated by minimizing the sum of squares between the log observed catch-per-unit-effort $\log(C_{a,y}/E_{a,y})$, and the log predicted catch-per-unit-effort $\log(\hat{C}_{a,y}/E_{a,y})$, i.e.,

$$SSR_{C/R} = \sum_{y=1}^{Y}\sum_{a=1}^{A}\Big[\log\Big(\frac{\hat{C}_{a,y}}{E_{a,y}}\Big) - \log\Big(\frac{C_{a,y}}{E_{a,y}}\Big)\Big]^2.$$

When catchability is not assumed constant, the estimation is performed as in Doubleday's method, described briefly in section 1.1.1.

Deriso et al. (1985) developed the first statistical catch-at-age software package that was based on the method of Doubleday (1976) but was more flexible in its options

for model formulations. Models were informed by auxiliary fishery-independent and -dependent data, including fishing effort and fecundity data, and there were options to select blocks of years and ages over which to estimate parameters. The package is no longer in use today, however two more up-to-date packages are described briefly below.

The age-structured assessment program (ASAP) was developed in the late 1990's and has a user-friendly interface that allows for easy model implementation (Legault & Restrepo 1998). The model is informed by total catch in weight, catch proportions in numbers, and auxiliary data via survey or commercial catch-per-unit-effort, which can all be specified per fleet (i.e. an aggregate of commercial fishing vessels). Model parameters are estimated in a maximum likelihood setting via AD Model Builder (ADMB; Fournier et al. 2012), a statistical application that uses automatic differentiation to provide the function optimizer with exact derivatives (Maunder et al. 2009). In recent years ASAP has mainly been applied in the management of US fisheries, including Atlantic mackerel (*Scomber scombrus*, NEFSC 2018) and Gulf of Maine cod (*Gadus morhua*, Palmer 2017).

The assessment method for Alaska (AMAK) is a software package that is informed by total landings, catch-at-age data, and research vessel survey biomass or abundance indices, although unlike many of the previously described software packages, AMAK can allow for sparse or missing data. Parameters are estimated in a likelihood setting via the ADMB package, however AMAK does not have a user-friendly interface and requires a working knowledge of ADMB and C++. AMAK is currently used for a variety of Northern stocks, including Aleutian Island pollock (*Gadus chalcogrammus*, Barbeaux & Paulson 2018) and the Atka mackerel stock in the Bering Sea and Aleutian Islands (*Pleurogrammus monopterygius*, Lowe & Paulson 2018).

Integrated models

Integrated models were introduced in the early 1980's by Fournier & Archibald (1982) and were the first attempt to incorporate errors in the model formulation that were not arbitrary. I present Fournier & Archibald (1982)'s original formulation however, as intended by the authors, it is very straightforward to fit variations of this model (e.g. changing the stock-recruit relationship or separability assumption). Fournier and Archibald began with a basic model that fit the landings data O_y and proportions at age $S_{a,y}$ separately in order to correctly account for errors associated with both data sources. The simplest model assumed $\sim \log(O_y) = \log(C_y) + \epsilon_O$, where C_y are the true unobserved landings, ϵ_O are assumed independent and normally distributed and $S_{a,y}$ are a random sample of the landings data with no ageing errors (observation equation is given below). There is not yet a link between the observed data and the underlying population processes, and Fournier and Archibald made the link via the assumption that catch at age are described by

$$C_{a,y} = C_y P_{a,y}, \tag{1.10}$$

where $P_{a,y}$ are the true unobserved proportion at age of fish in the catch. A separable fishing mortality is assumed (1.5), with s_a not estimated directly for each age, but by the parameters b_1, b_2 describing the shape of the selectivity curve. Fournier and Archibald argue that separability is more realistically described as a function of age,

$$s_a = b_1 a_w + b_2 a_w^2,$$

where the age index a is rescaled by

$$a_w = \frac{-1 + 2(1 - w^{a-1})}{(1 - w^{A-1})},$$

for $0 < w < 1$. Combining (1.10) and the catch-ratio model as in Doubleday (1976) gives

$$\log\left(\frac{C_y P_{a,y}}{C_{y+1} P_{a+1,y+1}}\right) = \log\left(\frac{F_{a,y} Z_{a+1,y+1}(1 - \exp^{-Z_{a,y}})}{\exp^{-Z_{a,y}} F_{a+1,y+1} Z_{a,y}(1 - \exp^{-Z_{a+1,y+1}})}\right) \tag{1.11}$$

Fournier and Archibald also incorporated auxiliary information in the model by making the assumption that yearly fishing mortality is proportional to fishing effort as in (1.7),

$$\log(F_{a,y}) = b_1 a_w + b_2 a_w^2 + \log(qE_y) + \epsilon_{Fy}, \tag{1.12}$$

where ϵ_{Fy} are assumed to be independent and normally distributed. Fecundity data can add additional information to the model, and are included via a Ricker spawner-recruit function,

$$\log(R_{r,y+r}) = \log(\alpha SP_y \exp^{-\gamma SP(y)}) + \epsilon_R y, \tag{1.13}$$

where SP_y is the reproductive potential of the population in year y, α and γ are parameters to be estimated and $\epsilon_F y$ are assumed to be independent and normally distributed.

Parameters are estimated in a maximum likelihood setting by minimizing the negative log-likelihood equation (nll), given by the corresponding contributions from the catch-at-age, fishing effort and fecundity data. The catch contribution to the nll

describes the catch proportions at age and landings (minus the constant terms), i.e.,

$$nll_C = \sum_a \sum_y S_{a,y} \log P_{a,y} - \sum_{i=1}^{Y} \frac{1}{2} \Big(\frac{\log O_y - \log C_y}{\sigma_c} \Big)^2 - n \log \sigma_c, \tag{1.14}$$

where n is the number of years of fishery observations. At the time of Fournier and Archibald's paper, this equation was impractical to solve and they proposed the simplifying transformation,

$$\beta_{a,y} = \log \Big[\frac{P_{a,y} C_y}{O_y} \Big].$$

It followed then that

$$\big(\log O_y - \log C_y \big)^2 = \log \Big(\sum_{a=1}^{A} \exp^{\beta_{a,y}} \Big)^2, \tag{1.15}$$

and

$$P_{a,y} = \frac{\exp^{\beta_{a,y}}}{\sum_{a=1}^{A} \exp^{\beta_{a,y}}}. \tag{1.16}$$

Substituting (1.15) and (1.16) into (1.14) gives the catch contribution to the nll,

$$nll_C = \sum_a \sum_y S_{a,y} \Big[\beta_{a,y} - \log \Big(\sum_{a=1}^{A} \exp^{\beta_{a,y}} \Big) - L_C \sum_{i=1}^{Y} \frac{1}{2} \log \Big(\sum_{a=1}^{A} \exp^{\beta_{a,y}} \Big)^2 \Big] - m \log \sigma_c,$$

where m is the greatest integer less than or equal to the number of age groups divided by two and $L_C = 1/\sigma_c^2$ is a penalty weight that determines the penalty for deviating from the observed catch relationship. These penalties weights are required because the age structure and effort data do not contain enough information to effectively estimate parameters (Fournier & Archibald 1982). The contribution to the nll from the effort data is

$$nll_F = -L_F \sum_{i=1}^{Y} \frac{1}{2} \epsilon_{yF}^2,$$

where similarly to above, $L_F = 1/\sigma_F^2$. The contribution from the fecundity data is,

$$nll_R = -L_R \sum_{i=1}^{Y-r} \frac{1}{2} \epsilon_{yR}^2,$$

with $L_R = 1/\sigma_R^2$. Thus, the full maximum likelihood equation is composed of three parts,

$$nll = nll_C + nll_F + nll_R.$$

It is straightforward to incorporate aging errors and/or additional data sources, and I refer the reader to Fournier & Archibald (1982) for more details.

A popular integrated stock assessment model software package (MULTIFAN-CL) was developed by Fournier et al. (1998, 1990) and can allow for size-based, age-structured, and even spatial-structured stock assessment model formulations. MULTIFAN is informed by commercial landings, length and/or weight proportions at age and fishery effort data and allows for region-specific recruitment, fleet-specific selectivity, and time-varying catchability (Hampton & Fournier 2001). Parameters are estimated in a "robust" likelihood setting to reduce the impact of outliers in the age composition data and are estimated via the ADMB package (see Fournier & Archibald 1982, for details). It has been noted that due to the complexity of the available model formulations, MULTIFAN is one of the more difficult stock assessment software packages to implement correctly (Dichmont et al. 2016). MULTIFAN is used today for many highly migratory fish stocks, including skip jack tuna (*Katsuwonus pelamis*) on the western and central Pacific ocean (Vincent & Hampton 2019).

In the early 2000's, a software package (CASAL) was designed to provide more consistent stock assessment advice for New Zealand fish stocks (Bull et al. 2005). CASAL model formulations can be size- or age-structured and can incorporate various types of data, including commercial catch-at-age, survey research biomass or

abundance indices and tagging data. CASAL is very flexible in its design and has provisions to account for multiple sub-stocks, multiple commercial fleets and can allow for the population to be structured by sex, maturity or growth. Parameters are estimated in a likelihood setting using a quasi-Newton optimizer from the ADOL-C project Betadiff package (Naumann & Schenk 2012). A majority of New Zealand fish stocks are managed using CASAL, including bluenose (*Hyperoglyphe antarctica*), gemfish (*Rexea solandri*) and black cardinal fish (*Epigonus telescopus*, FNZ 2019).

Stock Synthesis (SS) can fit generalized age- and length-based models (Methot Jr & Wetzel 2013) and unlike the software packages previously described, it is designed to not be critically dependent on one particular data type (ICES 2010). SS is very flexible and can range in application from data-limited situations to high-resolution spatial stock assessments. Parameters are estimated in a likelihood setting and are implemented via the ADMB package. SS has been used for stock assessments around the world, including short fin mako shark (*Isurus oxyrinchus*, ISC 2018b), blue shark (*Prionace glauca*) in the North Pacific Ocean (ISC 2017) and swordfish (*Xiphias gladius*) in the Western and Central North Pacific ocean (ISC 2018a).

State-space models

The theory underlying age-based state-space stock assessment models is not recent (Gudmundsson 1994, Sullivan 1992), however it was not until advances in computing power and technology that the application of state-space models was conceivable in practice as these models require integration over complex joint likelihood functions. In contrast to integrated models that treat the population processes as deterministic, state-space models allow for errors in the underlying population dynamics. There are two main components to a state-space model: the process model and the observation model.

The process model describes how the state of the unobserved fish stock abundance (i.e., the cohort equation (1.1)) and $F_{a,y}$ at a given time depend on previous states,

$$\alpha_y = T(\alpha_{y-1}) + \nu_y,$$

where α are the unobserved states and T is the transition function that describes the underlying population processes,

$$\begin{aligned} \log(N_{1,y}) &= \log(R_{1,y}) + \nu_{1,y}, \\ \log(N_{a,y}) &= \log(N_{a-1,y-1}) - Z_{a-1,y-1} + \nu_{a,y}, \end{aligned} \tag{1.17}$$

and $\nu_{a,y}$ are process errors that are typically assumed independent and normally distributed. Abundance at the earliest age $R_{1,y}$ can take on many forms (e.g. Ricker, Beverton-Holt), with the simplest being a random walk. Let $F_y = (F_{1,y}, F_{2,y}, ..., F_{A,y})$, then the fishing mortality rates are given by

$$\log(F_y) = \log(F_{y-1}) + \gamma_y,$$

where $\gamma_y \sim N(0, \Sigma)$ is the covariance matrix that can take on many forms (see, e.g. Nielsen & Berg 2014). In contrast to the previous methods, $F_{a,y}$ is now describing the unobserved process of fishing mortality, independent of $C_{a,y}$ that describes the actual catch.

The observation model describes how the catch-at-age data and the auxiliary data depend on the underlying population processes,

$$x_y = O(\alpha_y) + \epsilon_y.$$

Here, $O(\alpha_y)$ typically contains the catch equation (1.2) and an equation describing the research survey indices,

$$\log(I_{a,y}) = \log(q_{a,y}) + \log(N_{a,y}) + \epsilon_{Ia,y},$$
$$\log(C_{a,y}) = \log(\frac{F_{a,y}}{Z_{a,y}} N_{a,y}(1 - \exp^{-Z_{a,y}})) + \epsilon_{Ca,y},$$

where $\epsilon_{Ia,y}, \epsilon_{Ca,y}$ are the survey and catch errors, often assumed independent and normally distributed. The joint likelihood function of the unobserved states and observed states summarizes the state-space model,

$$L_{(\theta,\alpha,x)} = \prod_{y=2}^{Y} [\phi(\alpha_y - T(\alpha_{y-1}), \sum_{\nu})] \prod_{y=1}^{Y} [\phi(x_y - O(\alpha_y), \sum_{\epsilon})]$$

where θ are the parameters to be estimated and $\sum_{\nu}$, $\sum_{\epsilon}$ are the covariance matrices for the process and observation errors.

There are two approaches to conducting inferences via the joint likelihood function: the Bayesian approach and the frequentist approach. The frequentist approach considers θ as an unknown fixed parameter. The random effects are integrated out and estimation is performed via the marginal likelihood,

$$L_{M(\theta,x)} = \int L_{(\theta,\alpha,x)} d\alpha.$$

Although this integration was historically computationally difficult and time-consuming, the ADMB application implemented in R via the TMB package (Kristensen et al. 2015) can now efficiently integrate complex joint likelihood functions and has greatly increased the applicability of the frequentist approach in stock assessment modeling.

The Bayesian approach considers θ as a random variable, with prior knowledge about θ represented by $\pi(\theta)$,

$$p_{(\theta,x|\alpha)} = \frac{L_{(\theta,\alpha,x)}\pi(\theta)}{\iint L_{(\theta,\alpha,x)dxd\theta}}. \tag{1.18}$$

Simulation-based approximation methods are used to estimate model parameters, most commonly via variants of Markov chain Monte Carlo (see, e.g. Knape & De Valpine 2012, Team et al. 2016).

The state-space assessment model package (SAM) is currently the only software package that can fit state-space stock assessment models and is quickly become a favored tool for ICES stock assessments (Chouinard & Millar 2019). The model is informed by commercial catch-at-age and research survey indices, with assumptions for natural mortality, mean weights at age and proportion at age mature also required. The models can be fit and run online (Stockassessment.org) for certain stocks, or implemented via the R package "stockassessment". Stocks that are managed today via the SAM package, include North Sea, eastern English Channel and Skagerrak cod (*Gadus morhua*) and sole (*Solea solea*) in the Skagerrak, Kattegat, the Belts and Western Baltic sea (ICES 2019c).

Future of age-based stock assessment models

Most commercial fisheries that are managed using age-based stock assessment models ignore environmental effects, predation and the underlying spatial structure of the stock (Skern-Mauritzen et al. 2016). Historically, age-based models did not include these complexities as there was insufficient data and computational power with which to effectively estimate all parameters that more complex models required (Quinn 2003). However, advances in technology and data collection methods have

increased the availability of data to inform stock assessment models. The flexibility of state-space models is an ideal foundation from which to build more complex and realistic models and are increasingly the favored approach for stock assessment scientists (Aeberhard et al. 2018). Below, I describe the most recent advances in age-based state-space stock assessment models.

The definition of the ecosystem approach to fisheries management can vary wildly based on the management body and region (see, e.g., Ballesteros et al. 2017, Dolan et al. 2015), however all approaches are based on the simple underlying principle that ecosystem factors (e.g. species, environment) are intrinsically linked to other components (Harvey et al. 2016). When the interest is explicitly on fisheries with trophic interactions between the fish species, the term ecosystem based fisheries management (EBFM) is most often used (Dolan et al. 2015). EBFM has produced a wide range of models that range in complexity and ease of application. Whole ecosystem models that allow for interactions at all trophic levels have been developed (e.g. Ecosim, Christensen & Walters 2000), however these methods are simulation-based and more suited to the conceptual understanding of the ecosystem processes (Weijerman et al. 2016). Models that are an intermediate step between single-species models and full ecosystem models are better suited for management as they can estimate model parameters from the data (Plagányi et al. 2014). In the state-space framework, the effect of seal predation on cod has been incorporated into a single-species age-based model, with parameters estimated using the Bayesian approach (Cook et al. 2015, Holmes & Fryer 2011, Trijoulet et al. 2019, 2017). In the frequentist approach, Albertsen et al. (2017) improved single-species age-based stock assessment model fit by incorporating species correlation in cohort survival. These state-space models are a valuable first step to incorporating important ecosystem information in age-based stock assessment models.

The majority of the stock assessment methods described in the previous sections treat the stock as a spatially-aggregated closed population (i.e. no immigration or emigration within the management boundaries) with biological factors (e.g., mortality, growth) and fishing effort assumed homogeneous across the stock region. This is a major simplification and there is strong evidence that changes in the abundance of a fish stock can lead to major changes in the spatial distribution, and thereby availability of the fish to surveys and the commercial fishery (Ciannelli et al. 2013). Historically, it was often unfeasible or too expensive to collect spatially-explicit data, however, recent advances in technology and data collection methods have increased the availability and spatial resolution of stock assessment data. The last decade has also seen the implementation of spatially-driven management decisions, most notably marine protected areas (closures of specific areas to fishing) in many regions around the world (Berger et al., 2017). Marine protected areas can create spatial structure in fishing effort that cannot be accounted for by spatially-aggregated models (Field et al., 2006). Thus, although there is a greater availability of spatial stock assessment data and management decisions are more spatially driven than ever before, the transition to spatial stock assessment models has lagged. To the best of my knowledge, there are currently no spatial age-based state-space stock assessment models used for management in Canada and this should be a research topic of high priority if we are to ensure the sustainability of capture fisheries, especially those managed via spatially-derived decisions.

Almost all age-based stock assessment models are informed by pre-processed data. For example, catch-at-age data are derived from landings estimates from the various fishing fleets and sub-sampling of their catches for age and length measurements (see, e.g., Hilborn & Walters 2013) and commercial catch-per-unit-effort data are often standardized using generalized linear models (Gavaris 1988). The goal of integrated

stock assessment models is to correctly allow for the variability in all data sources in a likelihood framework. This approach often requires stock assessment modelers to make subjective decisions on the relative influence of each data set through data-weighting (see, e.g., Francis 2011, 2017). In "traditional" integrated models (i.e., not state-space) the weights aim to account for both measurement and process error, however in the state-space framework the process errors are explicitly accounted for (i.e. (1.17)). Thus, state-space models may provide an integrated framework that does not require subjective data-weighting since the process error is accounted for and measurement error can be included directly in the corresponding likelihood components. I do note that self-weighting in integrated state-space models requires further simulation and validation testing. Cadigan (2015) developed an integrated state-space stock assessment model for northern cod that did not require subjective data-weighting and is a promising step forward for integrated state-space stock assessment models.

1.2 Data and gaps

Throughout this text, I have touched on some of the complexities related to the most common fishery independent and dependent data sources (i.e., relative abundance indices, age/length composition and landings data) that inform age-based stock assessment models. Much of these data are not directly measured but are instead summary statistics. Historically almost all associated observation error was ignored or mis-specified (e.g., catch-at-age data in VPA), however newer model formulations can incorporate some data aggregation information (e.g. SAM can input standard errors for abundance indices).

Relative abundance indices

The standardization of research vessel and catch and effort data for input into stock assessment models has been well-studied. Historically, there were two main approaches to estimating relative abundance indices from both research vessel and catch and effort data: design-based estimators and model-based estimators. Design-based estimators estimate population abundance under the assumption that the randomness can be attributed to the sampling design, whereas model-based estimators assume that there is an underlying stochastic data-generating process (Smith 1990). Model-based estimators can account for underlying factors that design-based estimators cannot, e.g., differences in survey catchability (Helser et al. 2004) and time of the day and average depth of tow (Chen et al. 2004, Cotter & Pilling 2007). More recently, spatial models that assume that population densities at nearby sites are more similar than those that are further apart have been shown to produce more precise abundance indices than both the model- and design- based methods (Shelton et al. 2014, Thorson et al. 2016, 2015).

Commonly, only a subset of fish caught in a survey tow are aged as it is expensive and time-consuming to age every fish in the catch (Doubleday 1976). In this case, age-based indices are not independent in terms of sampling design and this can create complex correlations in the availability of the fish to the survey trawl. If all fish are measured for length it may be more appropriate to model separate likelihood functions for the length-based indices and age-length measurements combined with a stochastic growth model. If only a subset of the survey catch is measured for length and age then it may be better to model the aggregated survey indices separately from the length and age composition samples combined with a stochastic growth model. In both cases a stochastic growth model will be necessary in order to correct for the

length-selectivity of the survey trawl when estimating the distribution of length-at-age.

Landings data

Estimates of commercial landings can be derived from port-side sampling, fisher logbooks and observer records, however the precision and reliability of the estimates are rarely known (Cotter & Pilling 2007). The reliability of the landings estimates may vary throughout the time series for a multitude of reasons, including changes in observer coverage and/or mis-reporting from other countries. Unreported catch and discard rates also create discrepancies between estimated landings and actual at-sea removals, which can subsequently bias stock size estimates (Beare et al. 2005). Historically the landings data were not directly used in the stock assessment model or were treated as known (see, e.g., Francis 2017), however some state-space approaches have aimed to account for variability in the landings data. Cadigan (2015) used a censored-likelihood approach to account for underreporting in the landings data, and a formulation of the SAM software package can estimate a catch multiplier to allow estimation of bias in the reported landings (see, e.g., ICES 2011).

Composition data

Composition data represent the age or length distribution of the stock and are typically estimated from sub-sampling of research vessel or commercial data. These data are vital to age based stock assessment models and provide information on cohort strength and selectivity of the stock (Maunder & Piner 2014). In age-based assessment models catch-at-age data are a common source of composition data and are most often estimated from age-length keys (ALK) and length frequencies (LF). In a sampled commercial haul, all fish are measured for length, but only a subset are

selected to be aged, typically in a length-stratified age sampling design. The ALK can be calculated via the forward method (i.e. the probability of age given length, Fridriksson 1934), the inverse method (i.e. the probability of length given age, Clark 1981, Kimura & Chikuni 1987) or the mixed method (i.e. combination of both in a maximum likelihood framework, Hoenig et al. 2002). For brevity, I omit the details on the three methods but refer to Ailloud & Hoenig (2019) for an excellent description. The ALK is then applied to the unaged catch at length sample (LF) and the corresponding catch-at-age-by-length are summed over lengths to produce catch-at-age estimates (Ogle 2016).

Complex fishery sampling designs can lead to a lack of independence (i.e. over-dispersion) in the data due to size- or age-based fish behaviour (e.g. schooling, Pennington & Volstad 1994) and recent research has focused on how to best fit composition data within the stock assessment model. Francis (2014) argued that the commonly used multinomial likehood was poorly suited for compositionn data due to over-dispersion and concluded that the logistic-normal multinomial distribution showed great promise for composition data. Cadigan (2015) used a related approach, the multiplicative logistic-normal multinomial distribution, when fitting a state-space model for northern cod. More recently, Thorson et al. (2017) advocated for the Dirichelet-multinomial as an alternative to the multinomial, however Francis (2014) had previously noted the restrictiveness of the variance-covariance structure that only allows for negative correlations. Thus, how to best fit composition data within the stock assessment model remains a current topic of research.

1.3 Overview and Objectives

In this thesis, I address two of the data issues described above with a specific application to American plaice (*Hippoglossoides platessoides*) on the Grand Bank of

Newfoundland. A brief overview of the management of American plaice and available data is given below, followed by a discussion on the two issues that are addressed in this thesis.

American plaice

American plaice supported an important commercial fishery historically on the Grand Bank of Newfoundland (Northwest Atlantic Fisheries Organization (NAFO) Divisions 3LNO), at times accounting for over ten percent of Canadian groundfish landings (Morgan et al. 2011). The first stock assessment for American plaice was published in the early 1970's and set the total allowable catch (TAC) at 60,000 tons (Pitt, 1984). From 1973 to the late 1980's, the TAC ranged between 33,585 tons and 60,000 tons. The stock collapsed in the mid-nineties, and although the TAC has been set to zero tons since, the stock has yet to recover to pre-collapse levels. The current stock assessment model is an ADAPT-VPA that was introduced in the late 90's and is informed by survey abundance indices, catch-at-age data and biological data (Wheeland et al. 2018).

Relative abundance indices

There are currently three sources of fishery independent abundance indices for American plaice: the Canadian fall, Canadian spring and Spanish research vessel surveys. All surveys are stratified by depth and span different time periods: the fall survey has been conducted since 1981, the spring since 1971, and the Spanish survey since 1995. The spatial coverage varies across the three surveys and also within surveys. For example, the fall survey only covered the 3L region until 1990 when the 3NO regions were added. The Spanish survey covers the 3NO region, unlike the Canadian

surveys which covers 3LNO. For all surveys, the relative abundance indices are calculated from sub-samples of ages using a design-based method that accounts for the depth-stratified sampling design, (Smith & Somerton 1981).

Landings data

Landings data for American plaice have been collected since the early 1960's from both offshore and inshore fisheries via port and observer sampling (Stevenson 1983). There is considerable uncertainty about the landings data for American plaice in both the early and later years of the data (Wheeland et al. 2018). In the early years, sources of uncertainty include landings estimated for "unspecified flounder" by some countries (see, e.g., Pitt 1972) and an increase in foreign catch outside the 200 mile economic exclusive zone in the mid-80's (Brodie 1986). More recently, the loss of availability of scientific observer data in the NAFO Regulatory area has resulted in various methods being applied to obtain landings estimates, including effort ratios and daily catch records (Dwyer et al., 2016).

Composition data

Historically, catch-at-age data for American plaice were derived from stratification of the commercial landings (typically via area, time and gear) where the stratified sample was measured for weights and lengths and subsequent length-stratified subsamples were selected to be aged (Gavaris & Gavaris 1983). Since the 1994 closure of the directed commercial fishery various sources have been used to derive ages and lengths. From 1994-2000, ALKs were most commonly derived from the spring research survey vessels and applied to LFs from the Spanish and Portuguese fisheries. From 2001 onwards, ALKs were derived from the spring research survey vessels and

applied to LFs derived from the Canadian, Spanish and Portuguese fisheries by-catch of American plaice.

Objectives

This work is presented through two scientific papers that are detailed in the following chapters.

In Chapter 2, I present our first paper, where I improve our understanding of growth parameters through an extensive simulation study and application to American plaice. It was noted in Section 1.3 that when a subset of fish are caught in a survey tow to be aged then it may be more appropriate to model separate likelihood functions for the length-based indices and age-length measurements combined with a stochastic growth model. Aging only a subset of fish while all fish are measured for length is commonly referred to as length-stratified age sampling. In Chapter 2, I will improve our understanding and treatment of the estimation of growth parameters under this length-stratified age sampling design by collating and simulation testing nine methods found in the fisheries science and statistical literature to determine which method provides the most reliable growth model parameter estimates based on length-stratified age samples via simulated data. The optimal methods will then be applied to a case study for American plaice. This work will be a key component to fitting more realistic growth models that account for the length-stratified age sampling design and can be combined with length-based indices and age-length measurements to fit integrated stock assessment models with separate likelihood functions.

In Chapter 3, I present our second paper, where I develop an age-based state-space stock assessment model for American plaice that provides a better quantification of uncertainty. The current stock assessment model for American plaice is a virtual population analysis that treats the catch-at-age data as known without error. The

catch-at-age data are partially derived from landings estimates and there are considerable uncertainties about the landings data for this stock. I will develop an age-based state-space stock assessment model for Grand Bank American plaice that allows for uncertainties about the landings data and also in the underlying population processes (i.e., $N_{a,y}, F_{a,y}$). Model fit will be assessed via a detailed examination of model residuals for evidence of patterns indicating model misspecification and through retrospective model fitting. This work will provide a more realistic stock assessment model for American plaice on the Grand Bank of Newfoundland.

Chapter 2

Estimation of growth parameters based on length-stratified age samples

2.1 Introduction

Modeling and estimation of the relationship between fish length and age (i.e. growth) is an important component of fisheries science. Growth models are used in many ways, including the derivation of life history invariants (Charnov 1993, Jensen 1996), estimation of important population parameters such as selectivity and mortality (Hoggarth 2006, Hilborn & Walters 2013), and the classification of functional groups in ecosystems models (e.g. Shackell et al. 2010). Growth models are especially important in age-based fish stock assessments, where the estimates from these models are used in conjunction with a length-weight relationship in biomass calculations (Quinn & Deriso 1999), to convert length-based to age-based selectivity, and to estimate length compositions (Francis 2016). Stock assessment scientists must decide if growth will be estimated inside the stock assessment model via integrated stock assessment models (Maunder & Piner 2014, Methot Jr & Wetzel 2013), or outside the stock assessment model, where the fixed growth parameter estimates are used as data inputs for the assessment. Alternatively, empirical lengths and weights at age derived from individual fish measurements can also be applied instead of growth model estimates; however, this method requires high quality data across all assessment model years, which are often not available. Thus, reliable methods to estimate growth model parameters are

vital, as unreliable management decisions and conclusions can be drawn about the health of a fish stock if the growth estimates are incorrect (Smith et al. 1993).

The most widely used growth model in fisheries science is the von Bertalanffy model (vonB; Von Bertalanffy 1938), which assumes that the growth rate decreases linearly as size increases. The vonB parameters are estimated from samples of age and length measurements. Usually many more fish are measured for length than age because determining age is often time-consuming and expensive (Doubleday 1981) whereas lengths are much simpler to measure. A subsample of fish are selected for ageing. A popular sampling design for this is length-stratified age sampling (LSAS) which is a two-phase stratified sampling (TPSS) design. In the first sampling phase length is measured for the entire random sample of fish, and the fish are classified into length bins, or strata (e.g. one centimeter, five centimeters). In the second sampling phase, a small and prespecified number of fish are randomly selected to be aged from each length stratum. The rationale for this design is to reduce the number of ages to measure but maintain the representativeness of the data by ensuring enough age measurements in all the length intervals along the growth curve. This is an example of response-selective stratified sampling (RSSS) since length is the response and age is the descriptive covariate. There are many well-studied factors that can bias vonB parameter estimates, including selectivity, (e.g. Schueller et al. 2014, Taylor et al. 2005, Troynikov 1999), age measurement error (Cope & Punt 2007), and between-individual variability (see Kimura 2008, Sainsbury 1980, Vincenzi et al. 2016); however, bias from LSAS has not been sufficiently studied.

Although RSSS has been intensively studied, especially during the last two or three decades in the statistical literature (see, e.g. Breslow et al. 2009, Hsieh et al. 1985, Neyman 1938, Scott & Wild 2011), the application of the resulting statistical theories and methods to LSAS in fisheries science is inadequate. For example, popular

and easy-to-use approaches such as the weighted likelihood and calibrated weighted likelihood methods (detailed later; see Kalbfleisch & Lawless 1988, Saegusa & Wellner 2013) in RSSS are seldom applied in fisheries studies, and the few studies that do address LSAS have various theoretical or empirical issues to resolve. In the latter regard, Candy et al. (2007) accounted for LSAS with an approach for a different kind of RSSS, the variable probability sampling design, which still needs some theoretical and/or simulation validation. Another approach estimates mean length at each age, adjusted for LSAS through weights, called the bias-corrected mean length-at-age (BC) estimates (e.g. Bettoli & Miranda 2001). The BC estimates are treated as observed data and used to estimate growth model parameters using the nonlinear least squares method (e.g. see Brattey et al. 2018, Echave et al. 2012); however, the effect of using the BC estimates and not the individual observations on growth parameter estimates still requires further investigation. In addition, it is not possible to estimate between-individual variation in size-at-age with the BC approach, which is an important limitation of the method for some stock assessment models that require estimates of the stock distribution of size-at-age. Piner et al. (2016) and Lee et al. (2017) used length, instead of age, as the covariate by conditioning on length to circumvent the complexity with RSSS, since given length, the sub-sampling for aging is random in LSAS. However, here we show that this conditional method is both theoretically problematic and practically unreliable. On the other hand, theories and approaches proposed from statistical studies need some further development because they do not take into full account the specific complexities of fisheries surveys. For example, in LSAS, the number of length strata (100-200) is far more than that typical in the statistical literature on RSSS (about 10), and in LSAS, there are always some strata that are theoretically probable but practically empty, which is often neglected in statistical studies.

Zheng & Cadigan (2019) addressed some of these additional complexities in fisheries survey LSAS by proposing two new approaches. The first, a full-data likelihood function for LSAS, correctly accounted for the first phase data (i.e. only length), the fully observed second phase data (i.e. length and age), and the sampling scheme. The second approach improved on the method of Candy et al. (2007) which they referred to as the empirical proportion likelihood approach. Their simulation study indicated that these two new approaches perform as well as the standard full information likelihood approach, and better than the other existing approaches in RSSS. An important goal of this chapter is to formally compare the new approaches of Zheng & Cadigan (2019) with the existing methods for LSAS in the fisheries literature to determine one or two optimal LSAS vonB parameter estimation procedures. However in their simulation study, Zheng & Cadigan (2019) assumed that a simple parametric model for the age distribution of the population is known, which is usually not the case for fish stocks. Often the age distribution of a fish stock can be complex with multiple modes, and the age distribution changes from year to year as more abundant cohorts grow older. Another goal of this chapter is to address this deficiency.

Zheng & Cadigan (2019) concluded that for efficient and consistent parameter estimation when data are collected using LSAS, information from the following three components must be included: the first phase length data, the second phase length and age data (i.e. the subsample from phase 1), and the sampling scheme used to select fish for ageing. In this chapter we guide our estimating method comparison in light of these three components.We present the estimation performance of the vonB parameter estimates using nine estimation methods from the fisheries science and statistical literature. As a practical application, the two best performing methods were used to fit the vonB growth model for American plaice (*Hippoglossoides platessoides*) in Northwest Atlantic Fisheries Organization (NAFO) Divisions 3LNO.

2.2 Materials and Methods

Suppose we have N randomly selected fish with lengths and ages (l_i, a_i) for $i = 1, 2, ..., N$, generated from the distribution $f(l, a|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of all unknown parameters. The individuals are divided into K mutually exclusive and exhaustive length strata, $S_1, S_2..., S_K$. We measure all N fish for length in the first phase sample, but only a subsample of size n is measured for age in the second phase. Thus, the full age and length information (l_i, a_i) is only collected for a subset of the N fish. The probability that a fish with length l falls into the kth stratum (i.e. $l \in S_k, k = 1, 2, ..., K$) is given by

$$Q_k(\boldsymbol{\theta}) = P\{(l,a) \in S_k|\boldsymbol{\theta}\} = \int_{l\in S_k}\int_a f(l, a|\boldsymbol{\theta})\,\mathrm{d}a\,\mathrm{d}l. \tag{2.1}$$

We denote the number of fish in a length stratum by N_k, where $N = \sum_{k=1}^{K} N_k$. In each stratum, we specify a maximum number of fish to be sampled for age, m_k. The true sample size n_k for this stratum is

$$n_k = \begin{cases} N_k, & \text{if } N_k < m_k \\ m_k, & \text{if } N_k \geq m_k, \end{cases} \tag{2.2}$$

which is a random number. This is called length-stratified age sampling (LSAS).

In fisheries surveys, it is rare to sample the target m_k fish in strata for smaller and larger lengths. Fisheries survey gears may allow most smaller fish to escape, and it is uncommon to catch many fish of the oldest ages. As a result, the distribution of fish length is often highly right-skewed, and in this case the n_k's are random variables due to gear selectivity and mortality.

For LSAS (2.2), the density function for the second phase length-at-age data is

not $f(l, a|\boldsymbol{\theta})$ anymore, but given by (see Appendix A for derivation)

$$f(l, a|LSAS, \boldsymbol{\theta}) = \frac{f(l, a|\boldsymbol{\theta})}{Q_k(\boldsymbol{\theta})} Q_k(LSAS, \boldsymbol{\theta}), \tag{2.3}$$

where $Q_k(LSAS, \boldsymbol{\theta})$ is the probability for a second phase individual to be in length stratum S_k. Intuitively, $Q_k(LSAS, \boldsymbol{\theta})$ is related to the distribution of the random variable n_k. For instance, given n_k for each stratum, a second phase individual is twice as likely to be in a stratum with $n_k = 2$ than in a stratum with $n_k = 1$; if n_k was to change for a stratum, then $Q_k(LSAS, \boldsymbol{\theta})$ for the second phase stratum would also change. Zheng & Cadigan (2019) derived the exact formula for $Q_k(LSAS, \boldsymbol{\theta})$, which not only validates the above observation about the relationship between $Q_k(LSAS, \boldsymbol{\theta})$ and the distribution of n_k, but also indicates that $Q_k(LSAS, \boldsymbol{\theta})$ is indeed a function of the target parameters $\boldsymbol{\theta}$, and hence should not be neglected in $\boldsymbol{\theta}$ estimation. Their simulation study further confirmed that approaches neglecting $Q_k(LSAS, \boldsymbol{\theta})$ give fairly poor $\boldsymbol{\theta}$ estimates.

In the special case where the probability that $N_k < m_k$ is negligible for each stratum, then $n_k = m_k$ with probability 1 for each stratum. Namely, the n_k's are fixed numbers instead of random variables, and $Q_k(LSAS, \boldsymbol{\theta}) = \frac{m_k}{\sum m_k}$. This probability does not involve target parameters $\boldsymbol{\theta}$ and hence can be neglected for the purpose of $\boldsymbol{\theta}$ estimation. Some approaches (e.g. Hausman & Wise 1982) basically assume that the n_k's are always equal to m_k's, and hence neglect the $Q_k(LSAS, \boldsymbol{\theta})$ term in (2.3).

From the point of view of the three essential components for consistent and efficient parameter estimation, namely, the N first phase length data, the n second phase length-at-age data, and the sampling design, neglecting $Q_k(LSAS, \boldsymbol{\theta})$ is an incorrect incorporation of the sampling design, and hence leads to poorer estimation

results. Zheng & Cadigan (2019) suggested that all the deficiencies in RSSS estimation approaches to LSAS are due to insufficient or incorrect incorporation of one or more of these components. For the methods that we investigate in this chapter, the first phase information can be incorporated via the number of fish in each length bin (N_k) or the length of each fish. Incorporation of the sampling scheme indicates whether some measure has been taken to eliminate or alleviate the bias in the data introduced by LSAS. In the remainder of this chapter, we will categorize and analyze the existing LSAS estimation approaches in light of whether they incorporate these three components correctly and sufficiently.

Model and Likelihoods

Length-at-age is commonly described by the vonB growth model (e.g. Francis 2016, Quist et al. 2012), as it gives a good approximation of growth for many fish species (e.g. Chen et al. 1992, Quinn & Deriso 1999). The vonB model can be written as

$$l(a) = L_\infty\big(1 - e^{-k(a-a_0)}\big) + \epsilon, \tag{2.4}$$

where $l(a)$ is the length of an age a fish, L_∞ is the theoretical length at which the growth rate stops, k is the growth coefficient, and a_0 is the theoretical age at which length is zero. We assume the error term is given by $\epsilon \sim N(0, \sigma_a^2)$, where $\sigma_a = \mu_a\, CV$, CV is the coefficient of variation, and $\mu_a = L_\infty\big(1 - e^{-k(a-a_0)}\big)$. We assume this error term jointly represents the between-individual variation (or process error) resulting from individual growth parameter variability, plus the measurement error in length, as in Piner et al. (2016). Note that it is straight-froward to apply the estimation methods we investigate to other growth models (e.g. the Gompertz model).

Model (2.4) in fact states that the distribution of length l conditional on a,

Table 2.1: Log-likelihood functions used to estimate vonB population parameters, dbin represents the binomial probability mass function, pbin the cumulative binomial probability mass function, crl is the continuation ratio logit transformation, and equil. is the equilbrium approximation in equation 2.8.

Method	**Likelihood**	**P(a)**	**First phase**	**Sampling scheme**
Random	$\sum_l \sum_a n_{l,a} \cdot \log\Big(P(l\|a)P(a)\Big)$	Crl	No	No
Piner	$\sum_l \sum_a n_{l,a} \cdot \log\left(\frac{P(l\|a)\hat{P}(a)}{P(l)}\right)$	Equil.	No	Incorrect; neglects $\boldsymbol{\theta}$ in $P(l)$
PF	$\sum_l \sum_a n_{l,a} \cdot \log\left(\frac{P(l\|a)P(a)}{Q_k} Q_k(LSAS)\right)$	Crl	No	Yes
HT	$\sum_l \sum_a n_{l,a} \cdot \pi_{k(l)} \log\Big(P(l\|a)P(a)\Big)$	Crl	N_k	Yes, through weights
CW	$\sum_l \sum_a n_{l,a} \cdot w_i \log\Big(P(l\|a)P(a)\Big)$	Crl	N_k, l	Yes, through weights
Candy	$\sum_l \sum_a n_{l,a} \cdot \log\left(\frac{P(l\|a)}{\sum_k \frac{n_{k(l)}}{N_{k(l)}} Q'_k(a)}\right)$	n/a	N_k	Yes, ignores empty strata
EP	$\sum_l \sum_a n_{l,a} \cdot \log\left(\frac{P(l\|a)P(a)}{\sum_{k'=1}^{K_{obs}} \frac{n_{k(l)'}}{N_{k(l)'}} Q_{k'} + \sum_{k'=K_{obs}+1}^{K_{total}} Q_{k'}}\right)$	Crl	N_k	Yes, through EP density

Note BC and WBC methods are not listed as they are not likelihood approaches

$f(l|a) \sim N(\mu_a, \sigma_a^2)$. In practice, real length and age data are truncated to be integers, so the corresponding probability mass function (pmf) can be approximated as

$$P(l|a) = \frac{N(l, \mu_a, \sigma_a^2)}{\sum_l N(l, \mu_a, \sigma_a^2)}.$$

To fit five of the methods described below (see Table 2.1), the population distribution of age, $P(a)$, is required. In practice, the age distribution of a fish stock may be complicated and will change substantially over time because of the highly variable nature of fish recruitment. Simple models will usually not be reliable. In preliminary simulations (see Section 2.2) we tested many common probability distributions for $P(a)$, including the Poisson, Gamma and G-Normal mixture distributions as in Dey et al. (2019), but none of them performed well, as the randomly generated age distributions for our simulations (described in Section 2.2) varied from year to year. Hence, we decided to estimate the probability mass function *P(a)* at all ages directly.

We used the continuation ratio-logit transformation (e.g. Agresti 2003, Berg & Kristensen 2012, Cadigan 2015) which maps $P(a)$ for $a = 1, ..., A_{max}$ with the constraints $P(a) \geq 0$ and $\sum P(a) = 1$, into $\lambda(a)$ for $a = 1, ..., A_{max} - 1$ with no constraints, which are much easier to estimate. Thus,

$$P(a) = \begin{cases} \frac{\exp(\lambda(a))}{\prod_{i=1}^{a}\left(1+\exp(\lambda(i))\right)}, & a = 1, ..., A_{max} \\ \frac{1}{\prod_{i=1}^{A_{max}-1}\left(1+\exp(\lambda(i))\right)}, & a = A_{max}, \end{cases} \tag{2.5}$$

where A_{max} is the maximum age group for age groups $a = 1, ..., A_{max}$ and $\lambda(a) \in (-\infty, \infty)$ for $a = 1, \ldots, A_{max} - 1$. The inverse transformation of (3.6) is

$$\lambda(a) = \log\left[\frac{P(a)}{P(a+1) + ... + P(A_{max})}\right], \quad a = 1, ..., A_{max} - 1. \tag{2.6}$$

When implementing the various methods discussed in the following sections, $P(a)$ should be replaced by the functions of $\lambda(a)$'s in (3.6), and $\lambda(a)$'s are estimated together with the other model parameters.

Random method

Often in practice, because users are either not aware of, or don't understand the significance of, how their age-length data were collected, the LSAS design is inadvertently ignored and the second phase length and age sample are treated as a random sample from the population when estimating vonB parameters. We used this approach to demonstrate how biased and unreliable growth model parameter estimates can be obtained when this approach is applied to LSAS data. The most common way to implement this method fits the second phase length and age data to (2.4), and the vonB parameters are estimated using nls with age treated as a fixed covariate. The

random likelihood method described in Table 2.1 will give identical parameter estimates to the nls method when the distribution of age does not depend on the vonB parameters of interest. This method ignores both the first phase information and the sampling scheme, since it treats the second phase sample as a random sample.

Weighted methods

One of the most common approaches for statistical inference in RSSS is by weighing the likelihood function for the second phase data with information from the first phase sample (see Breslow et al. 2009, Kalbfleisch & Lawless 1988, Saegusa & Wellner 2013). These weighting methods are relatively easy to implement and can provide unbiased parameter estimating equations. The commonly used Horvitz-Thompson (HT) weight, $\pi_{k(l)} = \frac{N_{k(l)}}{n_{k(l)}}$, where $k(l)$ represents length l in stratum k, incorporates the first phase information and the sampling scheme through the fraction sampled for ageing ($\pi_{k(l)}^{-1}$; see Table 2.1 for all likelihood functions). Note that $N_{k(l)}$ contains some of the first phase length information. For example, if fish are measured by centimeter and one centimeter length bins are used for age sampling then $N_{k(l)}$'s provide all the first phase length information. If 5-cm length bins are used then the $N_{k(l)}$'s only provide partial first phase length information. The HT approach may not be efficient since the HT estimator may have large standard errors when the sampling units are not proportional to their inclusion probabilities (e.g. Thompson 2012). To adjust for this inefficiency, the calibrated-weighted (CW) method, first proposed by Deville & Särndal (1992), further adjusts the HT weights with first and/or higher order moments of the first phase length data (see Appendix B for details),

$$w_i = (1 - \lambda_c l_i)\frac{N_{k(l)}}{n_{k(l)}}. \tag{2.7}$$

Both the HT and CW methods incorporate some of the first phase information and the sampling scheme through their respective weights.

Conditional-on-length method

Piner et al. (2016) applied a conditional-on-length method based on the conditional probability

$$\begin{aligned} P(a|l) &= \frac{P(l \cap a)}{P(l)} \\ &= \frac{P(l|a)P(a)}{\sum_a P(l|a)P(a)}, \end{aligned}$$

to estimate the vonB parameters with data collected through LSAS. The rationale is that the conditional distribution of age given length does not depend on the length-stratified sampling design. The corresponding log-likelihood in Piner et al. (2016) is $\sum_l \sum_a n_{l,a} \log\big(P(a|l)\big)$. Note that $P(a,l) = P(a|l)P(l)$. The corresponding log-likelihood should be $\sum_l \sum_a n_{l,a} \log\big(P(a|l)\big) + \sum_l \sum_a n_{l,a} \log\big(P(l)\big)$. If $P(l)$ does not include model parameters $\boldsymbol{\theta}$ it can be neglected when applying the maximum likelihood approach. However, $P(l)$ as defined by Equation (2.4) must involve model parameters, and hence the second term involving $P(l)$ in the log-likelihood should be included. Therefore, the sampling scheme was incorrectly specified in the method of Piner et al. (2016), since it neglected the second term, $P(l)$. In addition, this method did not include any first phase length information. Piner et al. (2016) also used an equilibrium approximation for the distribution of ages,

$$\hat{P}(a) = \frac{e^{-Za}}{\sum_a e^{-Za}}, \tag{2.8}$$

where Z is the total mortality rate, defined as the sum of the natural mortality rate (M) and fishing mortality rate (F), Z=M+F. This approximation is based on an assumption of constant recruitment of age 1 fish. We note that this is the only method that requires an external estimate of M, as the other methods estimated $P(a)$ within the model (when required, see Table 2.1) using continuation ratio logits. This assumption for M is a large simplification as in reality natural mortality may not be time or location invariant. We investigate if the equilibrium approximation (2.8), which will often not be appropriate to describe the age distribution, is a useful assumption for estimating vonB parameters.

Conditional-on-age method

Candy et al. (2007) accounted for LSAS in vonB parameter estimation by treating the sampling scheme as variable probability sampling (VPS). VPS differs from LSAS in that instead of fixing the stratum subsample size, the sub-sampling probability of an individual is predetermined. For a discussion on the feasibility of using VPS for a sample collected through LSAS, we refer to Zheng & Cadigan (2019) and Candy et al. (2007). Candy's method differed from the traditional VPS estimation by substituting the observed proportions, $\frac{n_{k(l)}}{N_{k(l)}}$, which are independent of age, for the sampling probabilities. Candy et al. (2007) used the conditional probability given age to avoid the difficulty with the age distribution. The log-likelihood function for this approach is shown in Table 2.1. The conditional probability of being in the kth stratum is given by

$$Q'_k(a;\boldsymbol{\theta}) = \Pr\{l \in S_k | a\} = \int_{l \in S_k} f(l|a)\, dl. \tag{2.9}$$

Candy's approach incorporated the first phase information through the observed proportions, $\frac{n_{k(l)}}{N_{k(l)}}$. We consider conditioning on age further in the Discussion.

Improved conditional-on-age and full likelihood methods

Zheng & Cadigan (2019) proposed two likelihood approaches for parameter estimation in LSAS: the full-data likelihood method and the empirical proportion method (EP). Both the full-data and EP methods suggest incorporating the first phase data through the marginal density function of length, $f(l) = \sum_a f(l|a)P(a)$. However, the first phase density functions can only be correctly evaluated if there exists an appropriate age distribution for integrating a out of $f(l, a|\boldsymbol{\theta})$. In a preliminary analysis, we tried to infer the age-distribution from the LSAS data, however the initial results were not promising, and we did not pursue it further. Due to the lack of an effective method for estimating the age distribution for the simulated populations in this study, we estimated the vonB parameters using only the density functions for the second phase age and length data, which we call the EP method and the partial full-data (PF) method.

The EP approach improved on Candy's method by taking into account the scenarios where no fish are observed in a stratum (i.e. $N_k = 0$, see Table 2.1) and does not use the conditional-on-age approach (i.e. uses (2.1)). We discuss the drawbacks of using the probability of length conditional on age in the Discussion. The EP joint density function for l and a is

$$f_{EP}(l, a|LSAS; \boldsymbol{\theta}) = \frac{\frac{n_{k(l)}}{N_{k(l)}} f(l, a|\boldsymbol{\theta})}{\sum_{k'=1}^{K_{obs}} \frac{n_{k'(l)}}{N_{k'(l)}} Q_{k'} + \sum_{k'=K_{obs}+1}^{K_{total}} Q_{k'}}, \tag{2.10}$$

where $1, ..., K_{obs}$ are the strata with observed data, and $K_{obs} + 1, ..., K_{total}$ are the strata without data.

The PF method incorporates the sampling scheme correctly (see Zheng & Cadigan (2019) for full derivation) and the joint density function for l and a with $l \in S_k$ is

$$f_{\mathrm{PF}}(l,a|LSAS;\boldsymbol{\theta}) = \frac{f(l,a|\boldsymbol{\theta})}{Q_k} \times \frac{\left[\sum_{i=1}^{m_k-1} i \cdot \mathrm{dbin}(i,N,Q_k)\right] + m_k\left[1-\mathrm{pbin}(m_k-1,N,Q_k)\right]}{\sum_{k'=1}^{K}\left\{\left[\sum_{i=1}^{m_{k'}-1} i \cdot \mathrm{dbin}(i,N,Q_{k'})\right] + m_{k'}\left[1-\mathrm{pbin}(m_{k'}-1,N,Q_{k'})\right]\right\}}, \tag{2.11}$$

where dbin represents the binomial probability mass function, pbin the cumulative binomial probability mass function and m_k is the maximum sample size for full observation in stratum k. For the EP method, the first phase information was included through the empirical proportions and the sampling scheme was incorporated through the EP density function. The PF method correctly incorporated the sampling scheme, but only included first phase information via the total sample size N. Note that the total sample size provides very little information about the overall sampling scheme; for example the difference between $N = 1000$ in (2.11) and $N = 2000$ is negligible.

Mean length-at-age

We also studied fitting the vonB model to mean length-at-age (e.g. Echave et al. 2012) that were adjusted to account for bias due to the LSAS. This approach is commonly used in practice. The bias-corrected mean length-at-age (BC) estimate is

$$\bar{L}_a = \frac{\sum_k N_k(n_{a,k}/n_k)l_k}{\sum_k N_k(n_{a,k}/n_k)} \tag{2.12}$$

where l_k is the midpoint of length bin k, $n_{a,k}$ is the number of age a fish in length bin k, and N_k is the total number of fish sampled in length bin k in the first sampling phase. We use the BC estimates with (2.4), and estimate vonB growth model parameters using non-linear least squares (nls), both weighted by the number of fish at each age (WBC), and unweighted. This nls method includes both the sampling

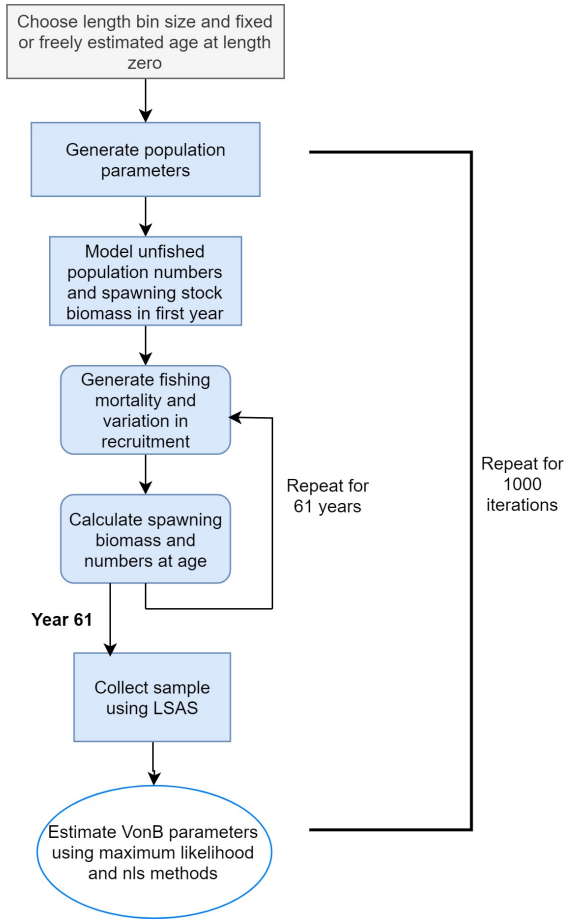
scheme and the first phase information through the weights in (2.12), but not the individual data. An important disadvantage of this approach is that estimates of σ_a^2 will not fully reflect between-individual variability.

Simulation study

Population simulations were similar to the approach described in Piner et al. (2016) where I: 1) generated a fish population using a wide range of life history parameters (see below), 2) simulated commercial fishery length-stratified age sampling, and 3) estimated growth parameters from the sample using the nine methods previously described (see Table 2.2 for details on distributions and models used in simulations). Simulations were designed to be typical of a fish stock and simulation settings were designed independently of the various growth model estimation methods I studied. Hence, the simulation procedure should not favor any of the estimation methods a priori. In this chapter, all the lengths are in centimeters and ages are in years. Recruitment for each year was modeled using the Beverton-Holt spawner-recruit model, parametrized in terms of steepness, h (e.g. Punt & Cope 2017). Length-at-age was simulated using a vonB model and weight at age was generated using a simple isometric growth model (Quinn & Deriso 1999). Maturity at age was assumed to follow a logistic model (e.g. Jennings et al. 2009), with age at 50% maturity, A_{50}, fixed at $\log(3)/k$, as in Jensen (1996), and age at 95% maturity A_{95} fixed at four years from A_{50}. The growth rate k was generated based on the life-history invariant $k = M/1.65$ (Charnov 1993). The simulation procedure is illustrated in Figure 1.

Starting from an unfished recruitment $R_0 = \exp(9)$, we simulated population dynamics for 61 years with these models and parameter specifications. For each year a

Figure 2.1: Diagram of steps used in population simulations; LSAS is length-stratified age sampling.



new fishing mortality and variation in recruitment were drawn from their corresponding distributions; all other population values were generated once per simulation run. The population in the 61st year was our first phase sample on which LSAS was conducted. The first phase sample size N was not fixed. Length bin sizes were chosen at one, two, three and five centimeters as they represented sizes that were frequently observed in practice (e.g. Monnahan et al. 2016). We estimated vonB growth parameters with length at age zero fixed at three centimeters, to compare with the results from Piner et al. (2016), and also estimated vonB parameters with a_0 estimated freely, as in practice this parameter is usually not fixed. For growth parameter

Table 2.2: Distributions and models used to generate population simulations; min = 0.10 and max = 0.50 for the uniform distribution; truncNorm is the truncated normal distribution with min = 0, max = ∞, to prevent generating negative values; when a_0 was estimated freely, a_0 was used, otherwise L_0 was used.

Parameter	**Distribution**	**Mean**	***SD***
Natural mortality (M)	Uniform	0.30	0.11
Fishing mortality (F)	TruncNorm	0.20	0.08
Steepness (h)	Normal	0.75	0.07
Log recruitment deviation (σ_{Rec})	TruncNorm	0.60	0.15
Asymptotic length (L_∞)	Normal	50	8
Growth coefficient error (ϵ)	Normal	0	$0.10(1.65/M)$
CV length-at-age	TruncNorm	0.10	0.04
Log unfished recruitment ($\log(R_0)$)	Fixed	9.0	-
Length at age 0 (L_0)	Fixed	3	-
Age at length 0 (a_0)	Fixed	-0.07	-
	Model		
Recruitment	$R(y) = \frac{SSB(y-1)}{1-\left(\frac{5h-1}{4h}\right)\left(1-\frac{SSB(y-1)}{SSB_{eq}}\right)}$		
Weight	$w(l(a)) = 0.20 \cdot l(a)^3$		
Maturity	$m(a) = \left(1 + e^{\frac{(-\log(19)(a-A_{50}))}{A_{95}-A_{50}}}\right)^{-1}$		
Numbers at age	$N(y,a) = N(y-1, a-1)e^{-(M+F(y))}$		
Numbers in first year	$N(1,a) = R_0\left(e^{-a \cdot M}\right)$		
Spawning stock biomass	$SSB(y) = \sum_{a=1}^{A} w(l(a)) \cdot m(a) \cdot N(y,a)$		

estimation using Piner's method, we estimated F as the median F across years within each simulation run, with M fixed at the true population value (see Table 2.2) for distributions used to derive population parameters), as this method performed well overall in Piner et al. (2016), and in practice F is often estimated from the data while M is fixed at a known value. All other methods did not require an estimate of M.

The purpose of LSAS is to (i) control the number of costly age measurements while (ii) obtaining length-at-age data evenly spread along the growth curve. Therefore, we assume that the target stratum sample size m_k is proportional to the length bin size so that the total target sample size $\sum_k m_k$ does not change with bin size (for

(i)), and that m_k is independent of k (for (ii), i.e. target m_k will be the same, no matter the distribution of length-at-age). To be specific, $m_k = 10 \cdot \text{binsize}$, for binsize $= 1, 2, 3, 5$. This choice of target stratum sample size is similar to sampling strategies used in real fisheries survey sampling (e.g. American plaice data used in Section 2.2). To demonstrate the effect of a change in the second phase sampling fraction (i.e. N and m_k) on the performance of the estimation approaches, we also conducted a simulation experiment with $m_k = 50$ and $R_0 = \exp(9)$, and with $m_k = 50$ and $R_0 = 5 \times \exp(9)$ for bin size equal to 1 cm and a_0 freely estimated.

This data generation and parameter estimation were repeated 1000 times. The estimation performance was measured using the relative root mean squared error (RRMSE),

$$RMSE = \sqrt{\frac{\sum_{i=1}^{1000}(\text{estimate}_i - \text{true}_i)^2}{1000}}, \qquad RRMSE = \frac{RMSE}{|\text{true*}|} \cdot 100,$$

and relative bias,

$$\text{Bias} = \frac{\sum_{i=1}^{1000}(\text{estimate}_i - \text{true}_i)}{1000}, \qquad \text{RelBias} = \frac{\text{Bias}}{|\text{true*}|} \cdot 100.$$

Here, true_i denotes the parameter value in the i^{th} simulation, and true* denotes the mean value of the parameter distribution. The parameters for L_∞, CV and k were randomly drawn from their corresponding distributions, and thus varied across iterations. For the life history invariant, k, 0.30/1.65 was used for true*, as 0.30 was the mean of the simulated natural mortality used in the calculation of k.

The estimation was performed in R using Template Model Builder (TMB) (Kristensen et al. 2015), an R package for fast evaluation of the negative log-likelihood

function and its gradients, which were then provided to the R function *nlminb*() for likelihood maximization.

American plaice growth

The two best fitting methods were used to estimate the vonB growth parameters using a dataset collected by the Department of Fisheries and Oceans Canada (DFO) in Northwest Atlantic Fisheries Organization (NAFO) Divisions 3L, 3N and 3O. We fit the data for the latest year available to us, 2014. The growth models were fit for female plaice only, since male and female plaice follow different growth curves. The actual sampling was conducted at various sites evenly spread across each division to ensure the spatial representativeness of the data, but with an overall sampling goal (m_k) in each division of about 25 age measurements per 2 cm length stratum by sex, if length was greater than or equal to 10 cm, and about 15 age measurements per stratum without distinguishing sex if length was less than 10 cm. For simplicity we assume that there is no spatial variation in size-at-age of these fish, and therefore can neglect the spatial structure of the sampling scheme and regard the whole sample in each division as a LSAS sample (see Appendix C).

The sampling in each division was treated as independent, so the likelihood function for the data is given by

$$L_{3LNO} = L_{3L} L_{3N} L_{3O}, \tag{2.13}$$

where L_{3L}, L_{3N} and L_{3O} are the likelihoods of each division. To compare the two best fitting methods with the methods that are commonly applied in practice, we also estimated vonB parameters using the BC method, both weighted and unweighted, and the random method.

2.3 Results

Results for simulation study

The random and Piner methods performed the worst overall. These methods ignored or misspecified both the sampling scheme and the first phase sample. When a_0 was fixed, Piner's method had the largest RRMSEs for all parameter estimates and for all length bin sizes (see Table 2.3). The RRMSE for k for Piner's method was much larger than the RRMSE for k for any other method when a_0 was fixed. For example, for the one centimeter length bin, the RRMSE for Piner's method was over 100%- all other methods had RRMSEs for k that were below 25% for that bin size. When a_0 was freely estimated, the RRMSEs for k using Piner's method were again much larger than the RRMSEs for k for all other methods (see Table 2.4). In terms of relative bias, Piner's method had the largest bias for k, no matter the bin size and whether or not a_0 was estimated freely. Overall, the relative bias of CV estimate was larger for the random method than for any other method.

The bias-corrected mean length-at-age methods, both weighted (WBC) and unweighted (BC), performed slightly better than the random and Piner methods, with RRMSEs that ranged between 5% and 23% for L_∞ and k, for all length bin sizes and whether or not a_0 was estimated freely. The RRMSEs for a_0 were always greater than 100%. The relative bias for the BC and WBC methods were larger than for any of the other methods that included the first phase length information and the sampling scheme (i.e. Horvitz-Thompson (HT), calibrated-weighted (CW), Candy and empirical proportion (EP)).

The partial-full (PF), HT, and CW methods were the next best fitting methods. The PF method had the smallest RRMSEs among all the methods that did not include the first phase sampling (i.e. Piner and random), with the RRMSEs decreasing

Table 2.3: RRMSEs (relative root mean squared error) and RelBias (relative bias) for vonB parameter estimates from 1000 simulations with length at age 0 fixed at 3 cm, for various length bin sizes. The abbreviations for the methods are: PF (partial-full), BC (bias-corrected mean length-at age), WBC (weighted BC), HT (Horvitz-Thompson weighted), CW (calibrated weighted) and EP (empirical proportion).

Method	**Random**	**Piner**	**PF**	**BC**	**WBC**	**HT**	**CW**	**Candy**	**EP**
RRMSE	1 centimeter								
L_∞	21.3	>100	3.5	11.0	5.2	2.7	2.7	1.7	1.6
k	23.1	>100	5.7	15.1	6.0	3.2	3.2	2.6	2.4
CV	27.0	34.5	3.9	-	-	5.0	5.0	4.1	3.9
	2 centimeters								
L_∞	21.6	>100	3.3	11.2	5.2	3.2	3.2	1.7	1.6
k	23.3	>100	5.7	15.3	6.2	4.0	4.0	2.6	2.5
CV	24.1	35.4	4.0	-	-	5.4	5.4	3.9	3.8
	3 centimeters								
L_∞	21.9	>100	3.3	11.0	5.3	3.6	3.6	1.7	1.6
k	23.6	>100	5.3	15.3	6.1	4.3	4.3	2.6	2.6
CV	21.5	33.9	4.1	-	-	5.7	5.7	3.8	3.8
	5 centimeters								
L_∞	23.2	>100	2.9	11.5	5.5	3.7	3.7	1.8	1.7
k	23.8	>100	4.8	15.9	7.2	4.7	4.7	2.8	2.8
CV	20.6	35.5	4.3	-	-	5.8	5.8	4.0	4.0
RelBias	1 centimeter								
L_∞	17.7	-0.5	0.9	7.9	3.3	0.0	0.0	0.1	-0.4
k	-18.9	>100	-1.2	-9.4	-1.5	0.1	0.1	-0.3	0.4
CV	25.7	-7.0	1.1	-	-	2.6	2.6	3.0	2.7
	2 centimeters								
L_∞	18.0	-4.0	0.5	8.1	3.3	0.0	0.0	0.0	-0.3
k	-19.1	>100	-0.5	-9.7	-1.5	0.2	0.2	-0.1	0.3
CV	21.9	-7.6	1.0	-	-	2.5	2.5	2.6	2.4
	3 centimeters								
L_∞	18.3	-7.5	0.5	8.1	3.4	0.2	0.2	-0.0	-0.3
k	-19.3	>100	-0.5	-9.8	-1.6	0.0	0.0	0.0	0.3
CV	18.4	-7.8	1.0	-	-	2.3	2.3	2.4	2.2
	5 centimeters								
L_∞	18.9	>100	0.2	8.6	3.6	0.1	0.2	-0.1	-0.3
k	-19.3	>100	-0.2	-10.3	-1.7	0.0	0.0	0.0	0.2
CV	14.3	-8.4	1.3	-	-	2.4	2.3	2.2	2.1

Note >100 are results that are greater than 100%

Table 2.4: RRMSEs (relative root mean squared error) and RelBias (relative bias) for vonB parameter estimates from 1000 simulations with a_0 estimated freely, for various length bin sizes. The abbreviations for the methods are: PF (partial-full), BC (bias-corrected mean length-at age), WBC (weighted BC), HT (Horvitz-Thompson weighted), CW (calibrated weighted) and EP (empirical proportion).

Method	**Random**	**Piner**	**PF**	**BC**	**WBC**	**HT**	**CW**	**Candy**	**EP**
RRMSE				1 centimeter					
L_∞	15.3	16.4	4.2	13.0	7.1	3.5	3.5	2.3	2.2
k	18.0	>100	11.3	21.9	10.9	5.4	5.4	4.4	4.2
CV	28.0	15.8	5.4	-	-	5.6	5.6	6.1	5.8
a_0	>100	>100	>100	>100	>100	43.7	43.7	39.4	37.2
				2 centimeters					
L_∞	14.8	>100	4.0	12.7	7.0	4.0	4.0	2.1	2.1
k	17.0	>100	9.5	22.0	10.9	5.8	5.8	4.2	4.2
CV	23.3	31.9	5.0	-	-	6.2	6.2	5.1	5.0
a_0	>100	>100	>100	>100	>100	56.7	56.7	46.2	44.6
				3 centimeters					
L_∞	14.2	16.3	3.7	12.9	7.2	4.8	4.8	2.1	2.1
k	15.7	>100	8.6	22.3	11.0	6.1	6.2	4.2	4.2
CV	20.1	16.3	4.4	-	-	6.2	6.2	4.6	4.5
a_0	>100	>100	92.0	>100	241.3	62.0	61.8	46.8	47.0
				5 centimeters					
L_∞	13.9	50.9	3.6	13.6	7.3	5.2	5.3	2.1	2.1
k	13.9	>100	8.1	23.0	12.1	7.1	7.1	4.5	4.5
CV	15.8	41.6	4.3	-	-	6.4	6.4	4.2	4.2
a_0	>100	>100	80.3	>100	>100	66.9	66.5	49.0	48.8
RelBias				1 centimeter					
L_∞	13.4	-1.3	0.3	8.9	5.1	0.4	0.4	0.0	-0.8
k	-14.6	>100	1.6	-13.7	-6.6	-0.4	-0.4	-0.5	0.9
CV	27.0	1.0	-0.1	-	-	3.6	3.6	4.4	4.0
a_0	10.3	>-100	39.4	>-100	>-100	-5.9	-5.9	-6.9	2.6
				2 centimeters					
L_∞	12.9	>100	0.5	9.0	5.1	0.5	0.5	-0.1	-0.6
k	-13.4	>100	0.2	-13.8	-6.6	-0.4	-0.4	-0.1	0.7
CV	20.8	2.4	0.8	-	-	3.4	3.4	3.4	3.1
a_0	38.9	>-100	11.3	>-100	>-100	-6.0	-6.2	-1.8	3.8
				3 centimeters					
L_∞	12.2	-1.0	0.5	9.1	5.1	0.5	0.4	-0.3	-0.6
k	-12.0	>100	-0.0	-14.0	-6.7	-0.3	-0.3	0.2	0.8
CV	16.4	0.4	1.2	-	-	3.2	3.2	3.0	2.8
a_0	65.1	>-100	3.7	>-100	>-100	-4.4	-4.1	2.4	6.3
				5 centimeters					
L_∞	11.6	1.2	0.3	9.5	5.2	0.7	0.7	-0.2	-0.4
k	-10.0	>100	0.1	-14.7	-6.7	-0.2	-0.2	0.3	0.7
CV	10.0	2.8	1.4	-	-	3.3	3.3	2.5	2.3
a_0	79.4	>-100	1.7	>-100	>-100	-4.5	-4.3	1.9	4.4

Note >±100 are results that are greater than 100%

as the length bin size increased. When a_0 was fixed, the HT and CW methods had identical RRMSEs, no matter the bin sizes - these RRMSEs increased slightly as the bin size increased, ranging between 3% and 6%. Overall, the relative bias when a_0 was fixed for both the HT and CW methods was small for all parameter estimates and length bin sizes. When a_0 was freely estimated, the CW and HT methods had almost identical RRMSE values for all parameter estimates and all length bin sizes, with the largest RRMSEs seen for a_0.

Table 2.5: RRMSEs (relative root mean squared error) and RelBias (relative bias) for vonB parameter estimates from 1000 simulations with a_0 estimated freely for 1cm bin sizes, target stratum sample size $m_k = 50$, and simulated population size (N) or five times simulated population size ($5N$). The abbreviations for the methods are: PF (partial-full), BC (bias-corrected mean length-at age), WBC (weighted BC), HT (Horvitz-Thompson weighted), CW (calibrated weighted) and EP (empirical proportion).

Method	**Random**	**Piner**	**PF**	**BC**	**WBC**	**HT**	**CW**	**Candy**	**EP**
RRMSE	N; 1 centimeter; m_k= 50								
L_∞	9.4	11.0	2.2	7.1	2.6	1.9	1.9	1.3	1.3
k	12.5	23.6	5.0	15.3	4.0	2.7	2.7	2.3	2.2
CV	24.3	7.5	3.2	-	-	5.0	5.0	5.4	5.3
a_0	>100	>100	68.0	>100	>100	29.9	29.9	25.5	25.0
RRMSE	$5N$; 1 centimeter; m_k= 50								
L_∞	14.9	11.1	2.1	9.4	3.3	1.7	1.7	1.2	1.3
k	18.1	25.6	5.4	20.8	5.5	2.6	2.6	2.0	2.0
CV	28.0	7.4	3.0	-	-	4.8	4.8	5.2	5.2
a_0	>100	>100	73.7	>100	>-100	27.8	27.8	22.1	21.9
RelBias	N; 1 centimeter; m_k= 50								
L_∞	7.9	0.8	0.2	3.5	1.7	0.2	0.2	-0.1	-0.5
k	-8.7	0.8	0.1	-4.8	-1.2	-0.3	-0.3	-0.2	0.4
CV	23.5	3.8	1.6	-	-	3.8	3.8	4.0	3.9
a_0	14.6	-31.1	11.1	>-100	>-100	-5.5	-5.5	-4.5	-0.6
RelBias	$5N$; 1 centimeter; m_k= 50								
L_∞	13.3	0.4	0.2	6.5	2.4	0.0	0.0	-0.5	-0.7
k	-14.9	2.2	-0.0	-12.8	-2.8	-0.1	-0.1	0.3	0.7
CV	27.2	2.8	1.2	-	-	3.5	3.5	4.0	4.0
a_0	18.5	-31.6	14.6	>-100	>-100	-2.5	-2.5	-0.6	1.9

Note >±100 are results that are greater than 100%

Out of all nine methods, EP and Candy's methods performed the best. When a_0 was fixed, the RRMSEs for the EP method were slightly smaller than the RRMSEs for Candy's method for all length bin sizes, with the difference in the RRMSEs decreasing as the bin size increased. Overall, the RRMSEs for both methods when a_0 was fixed were below 4%. When a_0 was freely estimated, the RRMSEs were smaller for all parameter estimates for the EP method for the one centimeter length bin, with the differences between the RRMSEs decreasing as the bin size increased. In terms of relative bias, Candy's method had slightly smaller relative bias for L_∞ and k than for the EP method, for all length bin sizes and whether or not a_0 was estimated freely. The relative bias for CV was slightly smaller for the EP method than for Candy's method, for all length bin sizes and whether or not a_0 was estimated freely.

In Table 2.5, when R_0 remained at exp(9), but m_k increased to 50 from 10 in Table 2.4, the performance of all the approaches including the random method improved with reduced RRMSEs and relative biases. When both R_0 and m_k increased by 5 times those in Table 2.4, namely R_0 increase to $5 \times \exp(9)$ and m_k to 50, the RRMSEs and relative biases of the random method did not improve, but the RRMSEs and relative bias of all the other approaches decreased.

Table 2.6: Parameter estimates and standard errors (SEs) for vonB model parameters for female American plaice data in NAFO Divisions 3LNO in 2014. The abbreviations for the methods are BC (bias-corrected mean length-at age), WBC (weighted BC) and EP (empirical proportion)

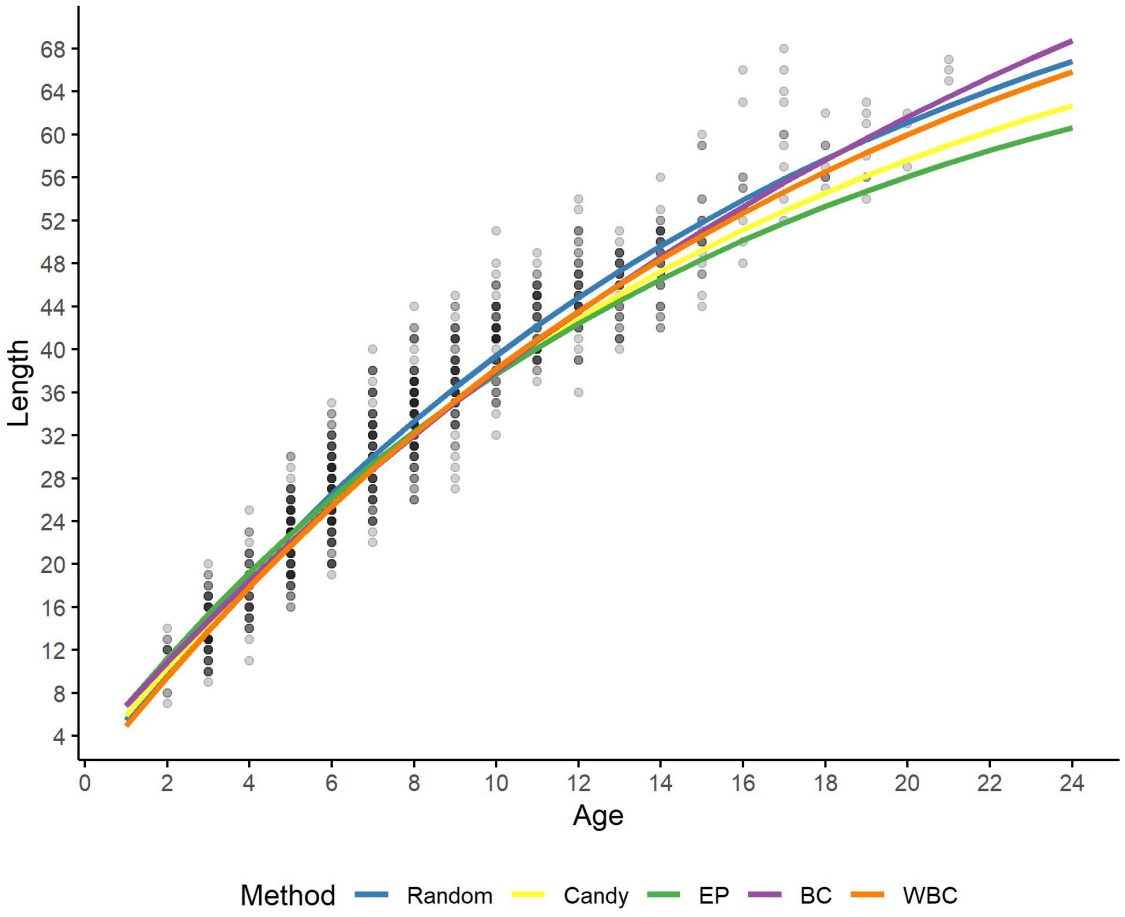
Method	L_∞	SE	k	SE	CV	SE	a_0	SE
Random	88.552	4.942	0.058	0.005	0.123	0.003	-0.097	0.092
BC	109.589	13.200	0.040	0.008	-	-	-0.603	0.364
WBC	89.472	7.620	0.055	0.008	-	-	-0.025	0.246
Candy	80.712	4.522	0.062	0.005	0.124	0.003	-0.223	0.106
EP	75.791	3.427	0.066	0.005	0.117	0.003	-0.425	0.112

American plaice growth

We estimated the vonB parameters for female American plaice in Division 3LNO in 2014 using the two best performing EP and Candy's methods. These methods produced smaller estimates for L_∞ and larger estimates for k than the random and BC methods (see Table 2.6). The EP method had the smallest standard error for L_∞, at 3.43. The standard errors for k and CV were identical for Candy's, EP and the random method. The BC methods had the largest estimates for L_∞, at 109.59 for the unweighted method, and 89.47 for the weighted method, with the largest standard errors overall for all parameter estimates.

In Figure 2.2 it may seem curious that the vonB growth curves estimated using Candy's and the EP methods do not capture the central trend in the data, and as such may seem biased. However, it is the LSAS data that are 'biased' in a random sample sense. Figure 2.3 shows the population (N) lengths-at-age for one of our simulated runs and Figure 2.4 shows the sub-sample (n) from that population. The predicted vonB growth curves using the second phase fully observed lengths and ages are also shown on each plot. It is clear from Figure 2.4 that the sub-sample is over-sampling larger older fish and shorter younger fish, which is not representative of the population. Figure 2.3 shows that the EP and Candy's methods are in fact best fitting as they capture the real trend in the population and not the LSAS bias in the sample. This LSAS bias in the data must be considered when assessing the validity of model assumption, and Zheng & Cadigan (2019) describe the method to compute residual diagnostics.

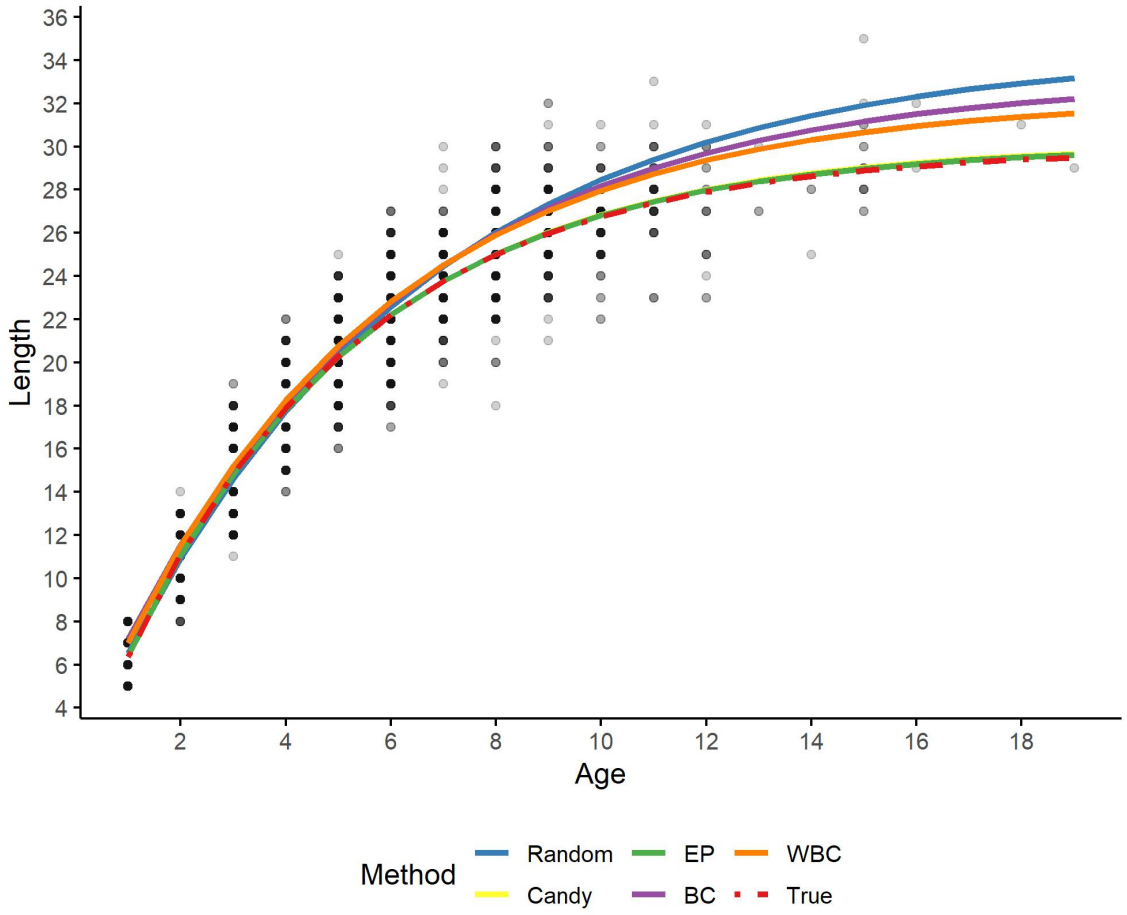
Figure 2.2: Estimated vonB growth model for female American plaice data in Divisions 3LNO in 2014. The points have added transparency to better represent the density of the observations. The various fits are: blue line (random), green line (EP), yellow line (Candy), purple line (bias-corrected mean length-at age (BC)) and orange line (WBC).



2.4 Discussion

Overall, our simulations showed that ignoring the first phase sampling information and misspecifying or ignoring the sampling scheme had large impacts on the accuracy of Von Bertalanffy (vonB) parameter estimates based on length-stratified age samples (LSAS). Both the Piner and random methods had large relative root mean squared

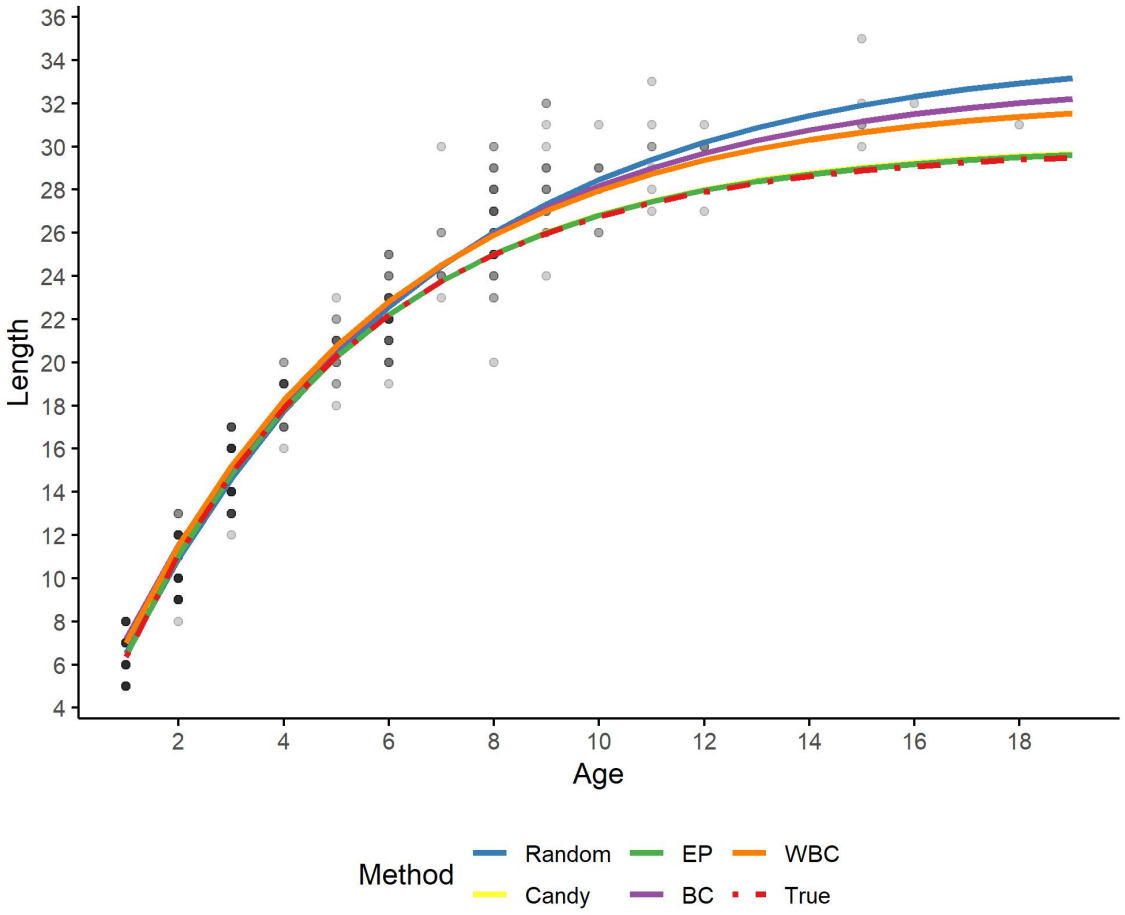
Figure 2.3: Plot of simulated population N with estimated vonB growth model from sub-sample n. The points have added transparency to better represent the density of the observations. The various fits are: blue line (random), green line (EP), yellow line (Candy), purple line (bias-corrected mean length-at age (BC)), orange line (WBC) and red line (population growth model)



error (RRMSE) for all length bin sizes. Incorporating the sampling scheme correctly, as in the partial-full (PF) method, reduced the RRMSE significantly.

The empirical proportion (EP) method performed better than the methods that did not include the first phase information (i.e. random, Piner), and the methods that used the bias-corrected mean length-at-age estimates both weighted (WBC) and unweighted (BC). The latter point is important because the bias-corrected methods

Figure 2.4: Plot of simulated sub-population n with estimated vonB growth model from sub-sample. The points have added transparency to better represent the density of the observations. The various fits are: blue line (random), green line (EP), yellow line (Candy), purple line (bias-corrected mean length-at age (BC)), orange line (WBC) and red line (population growth model).



are actively used in fisheries stock assessments to model growth (see, e.g. Brattey et al. 2018). The EP method also performed slightly better than Candy's method, and substantially better than the PF method that did not include first phase sampling information. In this simulation study the first phase samples were very large so that the strata with nonnegligible occupation probability Q_k were mostly nonempty, which reduced the improvement of the EP approach over Candy's method. It would be

interesting to investigate the magnitude of the difference in the RRMSEs when the first phase sample sizes are lower and as a result some of the length strata are empty because we expect the EP method will perform substantially better than Candy's approach in this situation. The full-data likelihood method proposed by Zheng and Cadigan (2019) seems less practically useful than Candy's or the EP method because of its performance and the strong assumptions (i.e. an effective estimation of age-distribution) required for the full-data likelihood method.

For almost all methods, when a_0 was freely estimated, the corresponding RRMSEs for a_0 increased as the bin size increased. This was not surprising because in our simulations the abundance of small fish sampled decreased with length bin size, so increasing the length bin size reduced the likelihood that a smaller fish was selected. Thus, for wider bins there were fewer fish sampled that were close to length zero, which provided less information with which to estimate a_0. As a result, when the bin size increased, the estimates of a_0 were less precise, which was reflected in the larger RRMSEs.

Theoretically, we expect Piner's approach to perform better than the random method, since it incorporated some information about the sampling scheme. Piner et al. (2016) also suggested that when the mortality rate Z was properly specified, their method performed better than the random approach for LSAS in terms of relative bias and the distribution of relative bias. Our simulations showed that the relative bias for Piner's method was consistently larger for k and for a_0. Our results also indicate that overall, Piner's method performed the worst in terms of RRMSE for all parameter estimates, length bin sizes, and whether or not a_0 was estimated freely, even though an approximately true mortality rate value was used for the Piner estimation. A primary reason for the poor performance of Piner's method was that it adopted an inappropriate model (2.8) for the age distribution. In comparison,

some of the other approaches, including the random method, used the estimated age distribution or conditioned on age, and hence were not influenced by this issue. Thus, mis-specifying the age distribution had a large impact on the RRMSE. Unless there is strong evidence for a specific age distribution for a population, we would not recommend specifying the age distribution outside the model.

The estimation methods in this chapter did not explicitly differentiate between individual variation in growth and length measurement error. All sources of error in the models were treated as one (i.e. confounded), and this is an idealized approach, because in reality there may be various sources of variability. Individuals may have L_∞ and k values that differ from the population mean parameters (Sainsbury 1980, Shelton et al. 2013). As well, age measurement error can occur in practice when conducting age readings on otoliths, and can lead to incorrect estimates of growth model parameters (e.g. Cope & Punt 2007, Dey et al. 2019). Not accounting for these sources of error can lead to biased estimates of L_∞ and k. Future research should assess the estimation performance of the methods in this chapter when age measurement errors are present.

We also note that although Candy's method performed well, there are two drawbacks to using the probability of length conditional on age. First, it is impossible to address age measurement errors with a structural errors in variables (SEV) approach (e.g. Carroll et al. 2006, Cope & Punt 2007, Dey et al. 2019). As noted above, the effect of ignoring age measurement errors is well known and can lead to incorrect estimates of growth parameters, and in particular under-estimation of L_∞ and over-estimation of k because of the well known bias-attenuation problem with covariate measurement errors. Such biases will tend to lead to over-estimation of F_{msy} and under-estimation of B_{msy}, and consequently suboptimal and possibly unsustainable harvest advice. Second, Candy's method cannot use the first phase length data with

the likelihood function as constructed in Zheng & Cadigan (2019). Due to these drawbacks, we suggest that the EP method may be easier to implement and more flexible in practice. Also due to these considerations, we tried to incorporate $P(a)$ in all the approaches except Candy, even though under the current simulation setup conditioning-on-age approach gives the same growth parameter estimates for random, HT and CW approaches since $P(a)$ involves no growth model parameters.

Increasing sample size leads to improved estimates if an estimator is unbiased. However the performance of a biased estimator may not improve with sample size. With LSAS, if the first phase stratum sample sizes $N_{k(l)}$ and target second phase stratum sample size m_k are all increased by 5 times for example, the bias in the data introduced by LSAS is not alleviated, and the performance of the estimation approaches neglecting LSAS do not improve. This is why when both R_0 and m_k increased by 5 times in Table 2.5 compared to Table 2.4, the performance of the random method did not improve as the other approaches did. Note that when the bin size equals 1cm, Piner's method is in fact the conditional approach discussed in Hausman & Wise (1982), Breslow & Cain (1988), Pfeffermann & Sverchkov (1999) and Scott & Wild (2011), which gives some account for the sampling strategy; hence, its performance also improved in this case. On the other hand, when the first phase sample size remained unchanged (equivalently R_0 remained at exp(9)), but m_k increased from 10 to 50, many strata became fully sampled or close to fully sampled for aging, the bias in LSAS data was alleviated, and the performance of random method also improved in Table 2.5. In the extreme case, when all first phase fish are selected for aging, then the data become a simple random sample from the population, and the random method will work well in this case.

The bias from mis-specifying the three components (i.e. first phase sample, second phase sample and sampling design) is not limited to data collected for growth parameter estimation. The stratified sampling design is often used to select sub-samples of fish to estimate important biological parameters such as length-at-maturity or length-fecundity relationships (Quinn & Deriso 1999). We suggest that the three components should not be neglected for any samples collected using stratified sampling and that parameter estimates will be biased if the sampling design is ignored. Further work should be done to investigate the magnitude of these biases for similarly collected data.

In conclusion, we demonstrate the importance of incorporating the first phase length information and the sampling design when estimating growth parameters with length-stratified age samples. Correctly accounting for the sampling scheme can greatly reduce the RRMSE for all parameter estimates. Incorporating the information from the first phase sample in combination with the sampling scheme reduced the RRMSE. We also showed that improperly modeling the age distribution can lead to large mis-specifications of growth parameter estimates, and using the LSAS bias corrected mean length-at-age can also produce less reliable parameter estimates. These growth parameter estimates are often incorporated into stock assessments to assess the overall health of a fish population. If the parameter estimates are incorrect, then we may be under or over-estimating the health of a fish stock, which can lead to incorrect management and stock advice.

2.5 Acknowledgments

Research funding to AP, NZ, and NC was provided by the Ocean Frontier Institute, through an award from the Canada First Research Excellence Fund. Research funding to NC was also provided by the Ocean Choice International Industry Research Chair

program at the Marine Institute of Memorial University of Newfoundland. Many thanks for the comments from the associate editor and two anonymous reviewers that greatly improved this manuscript.

Bibliography

Agresti, A. (2003). Categorical data analysis. Vol. 482. John Wiley & Sons.

Berg, C. W. & Kristensen, K. (2012). Spatial age-length key modelling using continuation ratio logits. Fisheries Research. **129**: 119–126.

Bettoli, P. W. & Miranda, L. E. (2001). Cautionary note about estimating mean length at age with subsampled data. North American Journal of Fisheries Management. **21**(2): 425–428.

Brattey, J., Cadigan, N., Dwyer, K., Healy, B., Ings, D., Lee, E., Maddock Parsons, D., Morgan, M., Regular, P. & Rideout, R. (2018). Assessment of the northern cod (*Gadus morhua*) stock in NAFO divisions 2J3KL in 2016. DFO Can. Sci. Advis. Sec. Res. Doc.. **2018/018**.

Breslow, N. E. & Cain, K. C. (1988). Logistic regression for two-stage case-control data. Biometrika. **75**(1): 11–20.

Breslow, N. E., Lumley, T., Ballantyne, C. M., Chambless, L. E. & Kulich, M. (2009). Improved Horvitz–Thompson estimation of model parameters from two-phase stratified samples: applications in epidemiology. Statistics in Biosciences. **1**(1): 32–49.

Cadigan, N. G. (2015). A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. Can. J. Fish. Aquat. Sci. **73**(2): 296–308.

Candy, S. G., Constable, A. J., Lamb, T. & Williams, R. (2007). A von Bertalanffy growth model for toothfish at Heard Island fitted to length-at-age data and compared to observed growth from mark–recapture studies. CCAMLR Science. **14**: 43–66.

Carroll, R. J., Ruppert, D., Stefanski, L. A. & Crainiceanu, C. M. (2006). Measurement error in nonlinear models: a modern perspective. CRC press.

Charnov, E. L. (1993). Life history invariants: some explorations of symmetry in evolutionary ecology. Vol. 6. Oxford University Press, USA.

Chen, Y., Jackson, D. & Harvey, H. (1992). A comparison of von Bertalanffy and polynomial functions in modelling fish growth data. Can. J. Fish. Aquat. Sci. **49**(6): 1228–1235.

Cope, J. M. & Punt, A. E. (2007). Admitting ageing error when fitting growth curves: an example using the von Bertalanffy growth function with random effects. Can. J. Fish. Aquat. Sci.. **64**(2): 205–218.

Deville, J.-C. & Särndal, C.-E. (1992). Calibration estimators in survey sampling. J. Am. Stat. Assoc. **87**(418): 376–382.

Dey, R., Cadigan, N. & Zheng, N. (2019). Estimation of the von Bertalanffy growth model when ages are measured with error. Journal of the Royal Statistical Society: Series C (Applied Statistics):In press. .

Doubleday, W. (1981). Bottom trawl surveys. Can Spec Pub Fish Aquat Sci. **58**: 1–273.

Echave, K. B., Hanselman, D. H., Adkison, M. D. & Sigler, M. F. (2012). Interdecadal change in growth of sablefish (*Anoplopoma fimbria*) in the Northeast Pacific ocean. Fish. Bull. **110**(3): 361–374.

Francis, R. C. (2016). Growth in age-structured stock assessment models. Fisheries Research. **180**: 77–86.

Hausman, J. A. & Wise, D. A. (1982). 'Stratification on endogenous variables and estimation: The gray income maintenance experiment'. in C. Manski & D. McFadden, eds, 'Structural Analysis of Discrete Data: With Econometric Applications'. M.I.T. Press. Cambridge, Mass. chapter 10, pp. 365–391.

Hilborn, R. & Walters, C. J. (2013). Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Springer Science & Business Media.

Hoggarth, D. D. (2006). Stock assessment for fishery management: A framework guide to the stock assessment tools of the fisheries management and science programme. number 487. Food & Agriculture Org.

Hsieh, D. A., Manski, C. F. & McFadden, D. (1985). Estimation of response probabilities from augmented retrospective observations. J. Am. Stat. Assoc. **80**(391): 651–662.

Jennings, S., Kaiser, M. & Reynolds, J. D. (2009). Marine fisheries ecology. John Wiley & Sons.

Jensen, A. (1996). Beverton and Holt life history invariants result from optimal trade-off of reproduction and survival. Can. J. Fish. Aquat. Sci. **53**(4): 820–822.

Kalbfleisch, J. D. & Lawless, J. F. (1988). Likelihood analysis of multi-state models for disease incidence and mortality. Statistics in Medicine. **7**(1-2): 149–160.

Kimura, D. K. (2008). Extending the von Bertalanffy growth model using explanatory variables. Can. J. Fish. Aquat. Sci. **65**(9): 1879–1891.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. & Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. Journal of Statistical Software. **070**(i05).

Lee, H., Thomas, L., Piner, K. & Maunder, M. (2017). Effects of age-based movement on the estimation of growth assuming random-at-age or random-at-length data. J. Fish Biol. **90**(1): 222–235.

Maunder, M. N. & Piner, K. R. (2014). Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. **72**(1): 7–18.

Methot, R. D. & Wetzel, C. R. (2013). Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research. **142**: 86–99.

Monnahan, C. C., Ono, K., Anderson, S. C., Rudd, M. B., Hicks, A. C., Hurtado-Ferro, F., Johnson, K. F., Kuriyama, P. T., Licandeo, R. R., Stawitz, C. C. et al. (2016). The effect of length bin width on growth estimation in integrated age-structured stock assessments. Fisheries Research. **180**: 103–112.

Neyman, J. (1938). Contribution to the theory of sampling human populations. J. Am. Stat. Assoc. **33**(201): 101–116.

Pfeffermann, D. & Sverchkov, M. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data. Sankhyā: The Indian Journal of Statistics, Series B. pp. 166–186.

Piner, K. R., Lee, H.-H. & Maunder, M. N. (2016). Evaluation of using random-at-length observations and an equilibrium approximation of the population age structure in fitting the von Bertalanffy growth function. Fisheries Research. **180**: 128–137.

Punt, A. E. & Cope, J. M. (2017). Extending integrated stock assessment models to use non-depensatory three-parameter stock-recruitment relationships. Fisheries research. .

Quinn, T. J. & Deriso, R. B. (1999). Quantitative fish dynamics. Oxford University Press.

Quist, M. C., Pegg, M. A. & DeVries, D. R. (2012). Age and growth. Fisheries techniques, 3rd edition. American Fisheries Society, Bethesda, Maryland. pp. 677–731.

Saegusa, T. & Wellner, J. A. (2013). Weighted likelihood estimation under two-phase sampling. Annals of statistics. **41**(1): 269.

Sainsbury, K. (1980). Effect of individual variability on the von Bertalanffy growth equation. Can. J. Fish. Aquat. Sci.. **37**(2): 241–247.

Schueller, A. M., Williams, E. H. & Cheshire, R. T. (2014). A proposed, tested, and applied adjustment to account for bias in growth parameter estimates due to selectivity. Fisheries Research. **158**: 26–39.

Scott, A. J. & Wild, C. J. (2011). Fitting regression models with response-biased samples. Canadian Journal of Statistics. **39**(3): 519–536.

Shackell, N. L., Frank, K. T., Fisher, J. A., Petrie, B. & Leggett, W. C. (2010). Decline in top predator body size and changing climate alter trophic structure in an oceanic ecosystem. Proc. R. Soc. London, Ser. B. **277**(1686): 1353–1360.

Shelton, A. O., Satterthwaite, W. H., Beakes, M. P., Munch, S. B., Sogard, S. M. & Mangel, M. (2013). Separating intrinsic and environmental contributions to growth and their population consequences. Am. Nat. **181**(6): 799–814.

Smith, S. J., Hunt, J. J. & Rivard, D. (1993). Risk Evaluation and Biological Reference Points for Fisheries Management Canadian Special Publication of Fisheries and Aquatic Sciences No. 120. NRC Research Press.

Taylor, N. G., Walters, C. J. & Martell, S. J. (2005). A new likelihood for simultaneously estimating von Bertalanffy growth parameters, gear selectivity, and natural and fishing mortality. Can. J. Fish. Aquat. Sci. **62**(1): 215–223.

Thompson, S. (2012). Sampling. John Wiley & Sons. Hoboken, N.J.

Troynikov, V. S. (1999). Use of Bayes theorem to correct size-specific sampling bias in growth data. Bulletin of Mathematical Biology. **61**(2): 355–363.

Vincenzi, S., Crivelli, A. J., Munch, S., Skaug, H. J. & Mangel, M. (2016). Trade-offs between accuracy and interpretability in von Bertalanffy random-effects models of growth. Ecological applications. **26**(5): 1535–1552.

Von Bertalanffy, L. (1938). A quantitative theory of organic growth (inquiries on growth laws. ii). Human biology. **10**(2): 181–213.

Zheng, N. & Cadigan, N. (2019). Likelihood methods for basic stratified sampling, with application to von bertalanffy growth model estimation. Journal of Statistical Planning and Inference. Submitted. .

Chapter 3

A state-space stock assessment model for American plaice on the Grand Bank of Newfoundland

3.1 Introduction

American plaice (*hippoglossoides platessoides*) on the Grand Bank of Newfoundland (NAFO Divisions 3LNO) supported an important commercial fishery historically, accounting for over ten percent of the Canadian groundfish fishery in the 1950's (Morgan et al. 2011). The population size declined rapidly in the 1980's due mostly to overfishing and, although there has been no directed commercial fishing since 1994, there has since been little improvement in the state of the population (see, e.g. Wheeland et al. 2018). The major factor that has been attributed to the lack of recovery is overfishing, which has occurred mainly through bycatch in the yellowtail flounder, skate, redfish, and Greenland halibut fisheries (Shelton & Morgan 2005). It has also been suggested that an increase in the natural mortality rate due to changing ocean temperatures may also be contributing to the lack of recovery (COSEWIC 2009).

The current stock assessment model for American plaice relies on the use of commercial catch-at-age data. When these data are available, the most commonly applied stock assessment models are virtual population analysis, statistical catch-at-age, and, more recently, integrated and/or state-space models. Virtual population analysis assumes that the catch-at-age data are known without error (Megrey 1989), whereas statistical catch-at-age models allow for errors in the catch-at-age data, although

these models commonly assume that the fisheries' selectivity pattern is constant over blocks of years (Hilborn & Walters 2013). Integrated models were first introduced in the early 1980's (Fournier & Archibald 1982) and aim to use as much data as possible in as raw a form as possible, while state-space models include both random errors in the underlying population dynamics model (i.e. for population abundance and fishing mortality rates) and measurement errors in the data (see, e.g. Albertsen et al. 2016, Cadigan 2015, Nielsen & Berg 2014). Advances in computing power and technology have led to an increase in the application of both integrated models (Maunder & Punt 2013) and state-space models (Aeberhard et al. 2018) in recent years as the availability and resolution of data has increased and it is now possible to efficiently integrate out random effects from complex joint likelihood functions.

The current stock assessment model for Grand Bank American plaice is a virtual population analysis that was introduced in the late 90's. This model is informed by catch-at-age data that are derived in part from landings estimates and does not account for the considerable uncertainty about the landings data (Wheeland et al. 2018). Sources of uncertainty include landings estimated from "unspecified flounder" by some countries in the earliest years of available data (see e.g. Pitt, 1972) and an increase in foreign catch outside the 200 mile economic exclusive zone in the mid-80's (e.g. Brodie, 1986). More recently, the loss of availability of scientific observer data in the NAFO Regulatory area has resulted in various methods applied to obtain landings estimates, including effort ratios and daily catch records (Dwyer et al., 2016). As a result, there may be large errors in the landings data and a stock assessment model that incorporates uncertainty in these data may therefore provide a better assessment of the stock.

Another issue that has been noted in the current assessment for American plaice are consistent directional changes in estimates of stock size as years of data are removed from the assessment model, called retrospective patterns (Mohn 1999). Retrospective patterns are caused by changes in the accuracy of the data over time and/or spatial and time-varying population processes that are unaccounted for or mis-specified in the model (see, e.g., Legault 2009). Systematic retrospective patterns can lead to incorrect management advice as important population processes (e.g. biomass and fishing mortality) may be over- or under-estimated and can result in unsustainable or sub-optimal harvesting advice (Szuwalski et al. 2017). To promote sustainable management advice for American plaice on the Grand Bank of Newfoundland, a stock assessment model that reduces or eliminates retrospective patterns is valuable. In this chapter, we develop an aged-based state-space stock assessment model for Grand Bank American plaice that allows for under-reporting in the landings data, accounts for the uncertainties in the catch-at-age data and reduces the problem of retrospective patterns.

3.2 Materials and Methods

There are two components to a state-space stock assessment model: the process model and the observation model. For our application to Grand Bank American plaice, the process model describes how the state of the unobserved fish stock abundance and fishing mortality rates at a given time depend on previous states. The observation model describes how the survey and commercial data depend on the unobserved states (see, e.g. Aeberhard et al. 2018).

Process model

The model runs for the years $y = 1960, ..., 2017$ for ages $a = 1, ..., 15^+$, where 15^+ represents the oldest ages grouped together from ages 15 onwards, called the plus group. For simplicity, we will refer to model ages $a = 1, ..., A^+$, and years $y = 1, ..., Y$. The process model describes how the abundance at age a in year y (i.e. $N_{y,a}$) and the fishing mortality, change over time. The $N_{y,a}$ for all ages and years are treated as random effects, with the cohort abundance model modelled as

$$\begin{aligned} \log(N_{y,a}) &= \log(N_{y-1,a-1}) - Z_{y-1,a-1} + \gamma_{y,a}, \\ \log(N_{y,A^+}) &= \log\left[N_{y-1,A^+-1}\exp^{-Z_{y-1,A^+-1}} + N_{y-1,A^+}\exp^{-Z_{y-1,A^+}}\right] + \gamma_{y,a}, \end{aligned} \tag{3.1}$$

where $Z_{y,a} = M_{y,a} + F_{y,a}$ is the total mortality rate given by the sum of the natural mortality rate, $M_{y,a}$ (i.e. all mortality unrelated to fishing) and F. Here, $M_{y,a}$ is assumed to be known and fixed at 0.50 for ages 1-3, 0.30 for age 4 and 0.20 for all ages 5 and above, except during 1989 to 1996, where it is fixed at 0.53 for all ages 5 and above, as recommended by Morgan & Brodie (2001), 0.83 for ages 1-3 and 0.63 for age 4. $F_{y,a}$ is set to zero for ages 1-4, as there is no reported catch at these ages. The $\gamma_{y,a}$ are the process errors, assumed to be independent and normally distributed with variance σ^2_{pe} to be estimated. The numbers at the first ages $N_{y,1}$ are modelled as

$$\log(N_{y,1}) = \mu_{R_y} + \delta_{R_y}, \tag{3.2}$$

where $\mu_{R_y} = \mu_{R_1}$ for $y \leq 1993$ and $\mu_{R_y} = \mu_{R_2}$ for $y > 1993$, and $\mu_{R_1}, \mu_{R_2} \in (-\infty, \infty)$, chosen to account for the large differences in recruitment between the two time periods and are fixed effect parameters to be estimated. The deviations from the mean recruitment δ_{R_y} are assumed to follow a normal distribution with AR(1) correlation across years, with the AR parameters σ^2_R and ϕ_R to be estimated, as we expect

recruitment to be more alike in years that are closer together. The fishing mortality rates are modelled similarly,

$$\log(F_{y,a}) = \mu_{F_{y,a}} + \delta_{F_{y,a}}, \tag{3.3}$$

where $\mu_{F_{y,a}}$ is the mean fishing mortality rate and $\delta_{F_{y,a}}$ is the deviation from the mean at each age and year. A separate $\mu_{F_{y,a}}$ is estimated for ages $5, 6, 7, 8, 9$ and 10^+ (where 10^+ represents ages $10 - 15^+$) for two blocks: 1960-1994 and 1995-2017 (i.e. twelve fixed effect F parameters). The age blocking of the $\mu_{F_{y,a}}$'s were chosen via model building to reflect overall fishery selectivity patterns, and the year blocks were chosen to account for the closure of the commercial fishery in 1994. The $\delta_{F_{y,a}}$'s are treated as random effects and are assumed to follow a normal distribution, with the deviations at the first age, $\delta_{F_{y,5}}$ assumed to have AR(1) correlation across years, independent from ages $6 - 15^+$, with parameters $\sigma^2_{F_5}, \phi_{F_5}$ to be estimated. We treat the $\delta_{F_{y,a}}$'s separately for age 5 fish as preliminary analyses indicated that trends in F's differed at age 5 compared to older ages. The F-deviations at ages $6 - 15^+$ were treated as a correlated AR(1) process across ages and years, with parameters $\sigma^2_{F6^+}, \phi_{F_{A6^+}}, \phi_{F_{Y6^+}}$ to be estimated. We fit an AR(1) process across ages and years for age $6 - 15^+$ fish as fish that are closer in age and time are expected to have F-deviations that are more similar than those that are further apart.

Observation model

The observation model includes data from the commercial fishery and scientific research trawl surveys. There are two basic types of fishery information: total landed weight, and the size (length, weight) and age composition of the landings. Both these sources of information are used to derive annual fishery catch numbers-at-age. In the

integrated assessment model philosophy, these data sources should enter into the assessment model fitting via separate observation models (i.e. one likelihood component for the age composition and one for the landings). We particularly want to focus our model estimation to include uncertainty in landings. Therefore, for pragmatic reasons, we used landings information (1960-2017) and the catch proportions-at-age (ages $5 - 15^{+}$ during 1960-2017) as independent data sources for model estimation. Stock size age-based indices are derived from the Canadian fall and spring research surveys in NAFO Divs. 3LNO (see Dwyer et al. 2016, for details) and the Spanish research survey in the portions of NAFO Divs. 3NO outside of the Canadian Exclusive Economic Zone (EEZ) (González-Troncoso et al. 2017) were also used in model estimation. Indices were for ages $1 - 15^{+}$ for all surveys, for years 1990-2017 for the fall survey (2004 and 2014 omitted due to poor survey coverage), 1985-2016 for the spring survey (2006 and 2015 omitted due to poor survey coverage) and 1997-2016 for the Spanish survey. The Baranov catch equation is used to model commercial catch as a function of N,F and Z,

$$C_{y,a} = \frac{F_{y,a}}{Z_{y,a}}(1 - \exp^{-Z_{y,a}})N_{y,a}. \tag{3.4}$$

Model predicted catch proportion at age ($P_a = C_a / \sum_a C_a$) were fit to observed proportions, as described in the next section. Commercial average weights-at-age ($W_{y,a}$) were calculated by Rivard's method (Rivard, 1980) and are used to calculate model predicted landings each year, $L_y = \sum_a W_{y,a} C_{y,a}$.

Age composition data

We fit the age composition data using the continuation-ratio logit (crl) transformation (see, e.g., Agresti 2003, Berg & Kristensen 2012, Cadigan 2015). A direct observation model for the matrix of observed catch proportions each year is complicated

because $P_{oa} \geq 0$ and $\sum P_{oa} = 1$. We use the crl which maps P_a for $a = 1, ..., A_{max}$ into $X_a \in (-\infty, \infty)$ for $a = 1, ..., A_{max} - 1$. The unconstrained crls are derived from the multiplicative logistic transformation,

$$X_a = \log\left[\frac{P_a}{P_{a+1} + ... + P_{A_{max}}}\right], a = 5, ..., A_{max} - 1. \tag{3.5}$$

where A_{max} is the plus group. The inverse transformation of (3.5) is

$$P(a) = \begin{cases} \frac{\exp(X(a))}{\prod_{i=1}^{a}\left(1+\exp(X_i)\right)}, & a = 5, ..., A_{max} - 1 \\ \frac{1}{\prod_{i=1}^{A_{max}-1}\left(1+\exp(X_i)\right)}, & a = A_{max}, \end{cases} \tag{3.6}$$

The crls for the observed catch proportions-at-age data (i.e. $X_{oy,a}$) are calculated from (3.5) and the observation equation (i.e. negative loglikelihood, nll) for the crls is

$$\text{nll}(X_{oy,a}|\boldsymbol{\theta}) = \sum_{y=1}^{Y}\sum_{a=1}^{A-1} \log\left[\phi\big(\frac{X_{oy,a} - X_{y,a}}{\sigma_{Ca}}\big)\right], \tag{3.7}$$

where ϕ is the probability distribution function (pdf) for a $N(0, 1)$ random variable and ϕ_{C_A}, ϕ_{C_Y} are AR(1) age and year correlation parameters to be estimated, as we expect the crl errors to be similar for fish that are closer in age and time. We fit two variance parameters ($\sigma^2_{C5}, \sigma^2_{C6+}$; one for age 5 fish, and one for ages 6+ fish) as preliminary data analysis indicated that the error variance differed at age 5 (see Fig. E.6), the youngest commercially caught age.

Landings data

Dwyer et al. (2016) reported uncertainties about the reliability of the landings data for Grand Bank American plaice. To account for this, we treat reported landings as a lower bound for true landings (i.e. not all catches are reported). We assume that

Table 3.1: Distributions and models used to generate population simulations; min = 0.10 and max = 0.50 for the uniform distribution; truncNorm is the truncated normal distribution with min = 0, max = ∞, to prevent generating negative values; when a_0 was estimated freely, a_0 was used, otherwise L_0 was used.

Period	UB	Comments
1960-1976	2xRC	"Unspecified flounder" by some countries, see, for example, (Pitt 1972)
1977-1982	1.2xRC	Landings by primarily Canada (>95%) after establishment of 200 mile EEZ
1983-1993	1.5xRC	increased foreign catch outside 200 miles; Various estimates used for catch; Issues with unspecified flounder records and discarding
1994-2010	1.2xRC	No directed fishing in 1994 (bycatch quota), 0 TAC 1995 onwards; Catches defined from various sources with those considered most reliable by Scientific Council used for totals
2011-2017	1.5xRC	Loss of availability of scientific observer data in the NAFO Regulatory area and surveillance estimates. Varying methods applies to obtain catch estimates including effort ratios (Dwyer et al. 2016), daily catch records, and NAFO CESAG estimates

there is an upper bound for landings that varies with the reliability of data (see Table 3.1 for details).

We assume the true landings could be accurately estimated with a CV of 2%. Let Bly and Buy denote the lower and upper bounds and $\sigma_C = 0.02$. The observation equation for the landings bounds data is

$$\text{nll}(L_1, \ldots, L_Y | \{B_{ly}, B_{uy}; y = 1, \ldots Y\}|) = \sum_{y=1}^{Y} \log \left[\Phi \left\{ \frac{\log B_{uy}/L_y)}{\sigma_C} \right\} - \Phi \left\{ \frac{\log(B_{ly}/L_y)}{\sigma_C} \right\} \right] \tag{3.8}$$

where $L_1, \ldots, L_Y$ are the model predicted landings. We fixed σ_C at 0.02 to ensure that the estimates of landings are between the bounds for most years.

Survey data

The model-predicted catch for survey s is

$$I_{s,y,a} = q_{s,a} N_{y,a} \exp^{-f_{s,y} Z_{y,a}} \tag{3.9}$$

where f represents the fraction of the year the survey takes place (0.460 for the Canadian spring and Spanish surveys and 0.875 for the Canadian fall survey), and $q_{s,a}$ are the survey catchabilities that are free parameters to be estimated. The q parameters are fit by survey with a separate q estimated for ages 1-9 and grouped for ages 10^{+}, as preliminary model fitting resulted in similar q parameters at the oldest ages and this reduced the number of parameters to be estimated. For the spring and fall surveys, fish of ages 1-4 are given a separate q for each gear period due to issues in conversion from the Engel to the Campelen survey trawl (Dwyer et al. 2016). Diagnostic model fitting found little difference in the q estimates for the Spanish survey for the two gear types at the youngest ages, therefore only one q was estimated at each age. The indices are assumed to follow a normal distribution with mean $I_{s,y,a}$ and standard deviation $\sigma_{s,a} = \text{cv}_s \cdot I_{s,y,a}$, where cv_s represents a separate coefficient of variation (CV) parameter for each survey, to be estimated. The cv_s parameters are pooled by ages 1, 2-5, 6-9, and 10-15 for the fall and Spanish surveys, and by ages 1, 2, 3-5, 6-9, and 10-15 for the spring survey, as diagnostic model runs suggested differing coefficients of variation at these ages (see Fig. F.5). The observation equation for the survey data is

$$\text{nll}(I_{s,y,a}|\boldsymbol{\theta}) = \sum_{y}\sum_{a} \log\left[\phi\Big(\frac{I_{s,y,a} - EI_{s,y,a}}{\sigma_{s,a}}\Big)\right]. \tag{3.10}$$

We treated each survey as from an AR(1) process across ages with independent parameters ϕ_s to be estimated. A constant CV variance model for I is approximately the same as assuming $\log(I)$ has constant variance; however, an advantage of our approach is that we can use observed zero indices directly in the model whereas in other assessment packages these index zeros are typically excluded which is not appropriate when there are many zeros.

Table 3.2: State-space model fixed effect parameters

Type	Parameters
Process error variance	σ^2_{pe}
Mean recruitment	μ_{R_1} for $y \leq 1993$, μ_{R_2} for $y > 1993$
Mean $\log_F$	$\mu_{F_{y,a}}$ for ages 5, 6, 7, 8, 9, 10+; two blocks 1960-1994/1995-2017
Variance and year-correlation of $\log_F$ devs at age 5	$\sigma^2_{F_5}, \phi_{F_5}$
Variance, age- and year-correlation of $\log_F$ devs at ages 6-15+	$\sigma^2_{F6^+}, \phi_{F_{A6^+}}, \phi_{F_{Y6+}}$
Catch age composition variance and correlation	$\sigma^2_{C5}, \sigma^2_{C6+}, \phi_{C_A}, \phi_{C_Y}$
Canadian Spring Survey (s)	$q_{s,1E}, \ldots, q_{s,4E}, q_{s,1C}, \ldots, q_{s,4C}, q_{s,5}, \ldots, q_{s,9}, q_{s,10+}$
Canadian Fall Survey (f)	$q_{f,1E}, \ldots, q_{f,4E}, q_{f,1C}, \ldots, q_{f,4C}, q_{f,5}, \ldots, q_{f,9}, q_{f,10+}$
Spanish Survey (ss)	$q_{ss,1}, \ldots, q_{ss,9}, q_{ss,10+}$
Canadian Spring Survey CV	grouped ages 1, 2, 3-5, 6-9, 10-15
Canadian Fall Survey CV	grouped ages 1, 2-5, 6-9, 10-15
Spanish Survey CV	grouped ages 1, 2-5, 6-9, 10-15
Survey residual correlations	for s = Canadian spring, fall Spanish surveys

Estimation

The fixed-effect parameters to estimate (i.e. θ) are listed in Table 3.2. The unobserved states (i.e. $\delta_{F_{y,a}}, N_{y,a}$) are integrated out of the joint likelihood functions and the estimation θ is based on maximizing the marginal likelihood $L(\theta)$,

$$L(\theta) = \iiint_{\Psi} f_\theta(D|\Psi) g_\theta(\Psi) \partial\Psi \tag{3.11}$$

where Ψ is the vector of all random effects, $f_\theta(D|\Psi)$ is the joint probability density function of the data (i.e. $\log f_\theta(D|\Psi)$ is the sum of the observation equations 3.7, 3.8 and 3.10) and $g_\theta(\Psi)$ is the joint probability density function for the random effects. The TMB (Kristensen et al. 2015) package in R is used to integrate the marginal likelihood (1.8), which is performed via the Laplace approximation (see Skaug & Fournier 2006, for details). The nlminb package in R is used to minimize the negative log likelihood function provided by TMB.

The model fit was assessed by examining the model residuals and retrospective plots. The survey and continuation-ratio logit residual observation model were adjusted to account for observed variance heterogeneity and correlation during model

building runs. Retrospective models were fit for years 2005-2017, with each retrospective model fit using one less year of data (i.e. model for year 2005 used data up to 2005) and predicted abundance, biomass, spawning stock biomass and average F's were plotted and examined for systematic patterns. Ideally, no discernable directional patterns will be present in the retrospective plots.

Biomass-at-age was calculated by multiplying predicted numbers at age (i.e. $N_{y,a}$) and stock weights-at-age, which were estimated externally. A spatiotemporal biphasic von Bertalanffy growth model (see Kumar et al., 2019) was applied to length-at-age data obtained from the spring research survey data. The combined 3LNO stock weights were estimated by weighting the stock weights for each division by the average index at age during 1975-2017. Stock weights prior to 1975 were fixed at the mean values for 1975-77. Estimates of maturity-at-age were taken from Wheeland et al. (2018).

Simulation and sensitivity testing

A full simulation study is beyond the scope of this chapter; however, we conducted a simple self-simulation test and jittered start test to examine the reliability of the model estimates (Cadigan 2015, Nielsen & Berg 2014). The self-simulation test randomly generates survey indices and continuation-ratio logit catch proportions from the model predictions and assumed distributions detailed above. Process errors and random effects are treated as fixed and the model is re-fitted to the simulated data. This process is repeated 1000 times and estimates of SSB, average fishing mortality rates (ages 9-14) and recruitment are stored. We calculated the relative difference of the estimates for each year (i.e. (simulation SSBy – data-based SSBy)/ data-based SSBy) for comparison. The jittered start test re-fits the model with random noise added to the starting parameter values, generated from $N(0, 0.25\hat{\mu})$, where $\hat{\mu}$ is the

model predicted parameter of interest. The model is re-optimized 50 times and the negative log-likelihood is stored for each iteration. Ideally, we expect an identical model fit from the jittered starting parameter values. We also examined the model sensitivity to our assumptions about M, upper catch bounds and survey catchability, q. We re-fit the model with M's fixed at 0.20 for all ages and years (M2), M's fixed at the original model formulation plus 0.20 (M3), with upper catch bounds fixed at half the original model formulation upper bounds (M4), and with q estimated freely at the oldest ages (M5).

3.3 Results

The state space model (SSM) fit the data well with no patterns present in the survey or continuation-ratio logit residual plots (see Appendix D). In 2017 recruitment, abundance and spawning stock biomass (SSB) were estimated near the lowest historical levels (Fig. 3.1). The model predicted landings were estimated within the upper and lower bounds, with the predicted landings closest to the upper bound in the early 80's, and again in most years since 2006 (Fig. 3.2) and closest to the lower bound in the early 1990's. At ages 1-4, the catchability pattern (Fig. 3.3) for the fall and spring surveys was lower for the Engels than the Campelen trawl. The differences were most pronounced for ages three and four, with the catchability estimates for the Campelen trawl almost twice as large as for the Engels trawl. For ages 1-5, the process errors (Fig. 3.4) were close to zero until the mid-nineties. The process errors were similar for ages 7-9 and again for ages 13-15. There were little retrospective patterns (Fig. 3.5), with SSB slightly overestimated and average fishing mortality for ages 9-14 (aveF) slightly underestimated.

In comparison to the most recent stock assessment model for Grand Bank American plaice (which we refer to as the VPA), the overall trends in SSB and aveF were

Figure 3.1: Model estimated population abundance, spawning stock biomass, average fishing mortality rates (ages 9-14) for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent 95% confidence intervals.

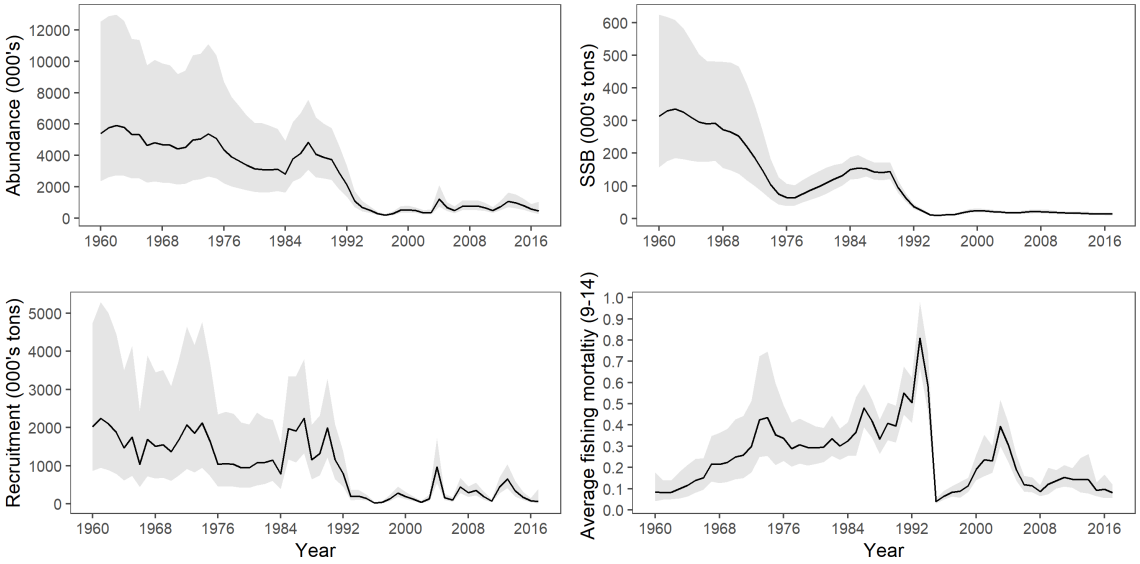
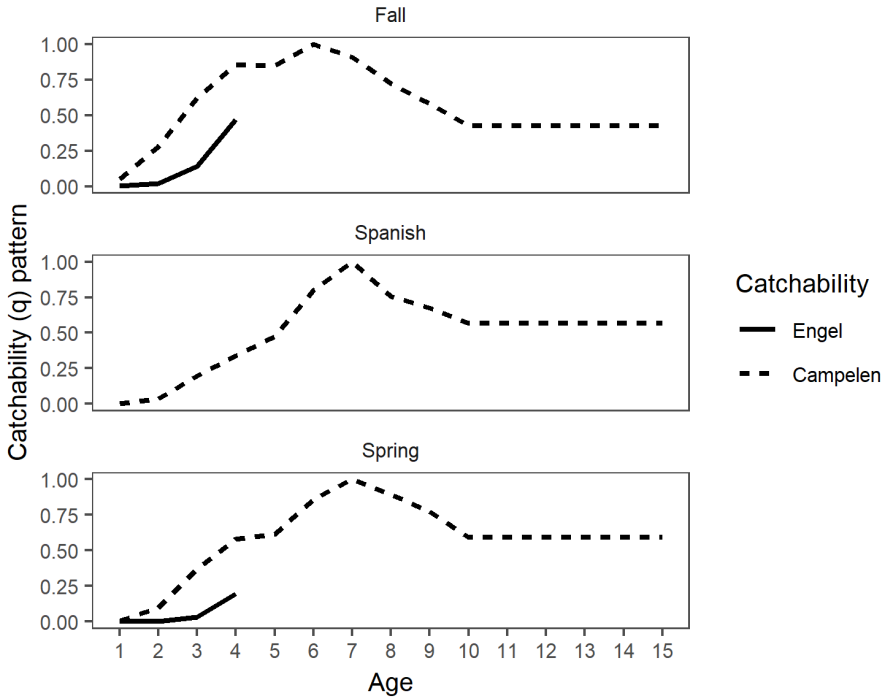


Figure 3.2: Model estimated log catch numbers for ages 5-15+ (solid line) for American plaice in NAFO Divisions 3LNO, the shaded grey represents the region between the log lower catch bounds and the log upper catch bounds.

Figure 3.3: Survey catchability patterns for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. A separate catchability parameter is estimated for two gear types (Engel and Campelen) for the spring and fall surveys for ages 1-4.

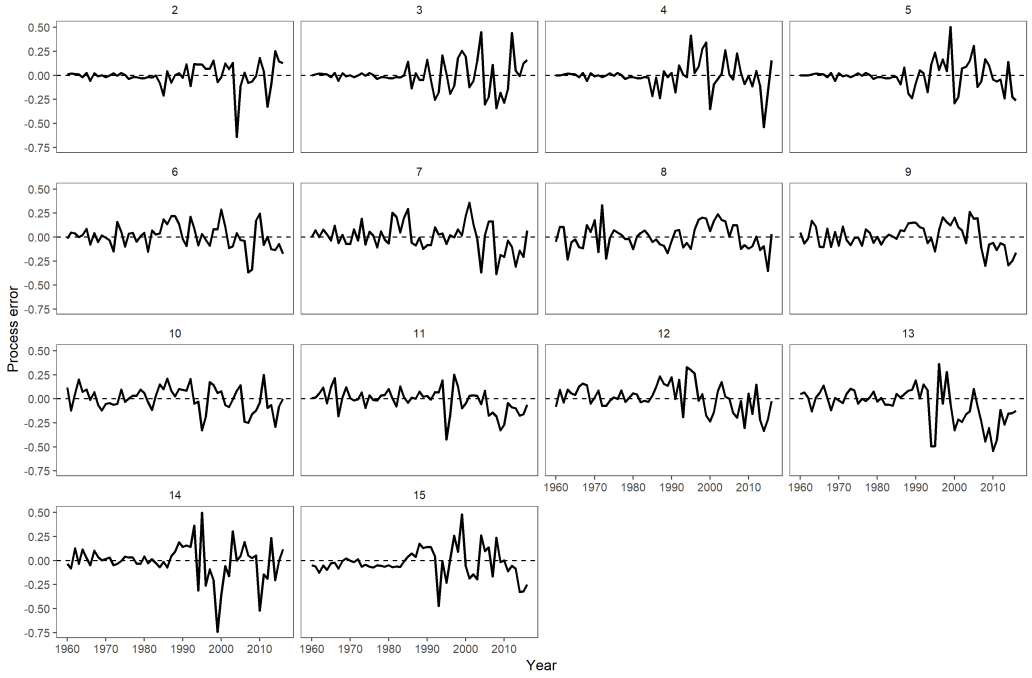


similar (Fig. 3.6 and Fig. 3.7). Noticeable differences included the SSM predictions of historic SSB (i.e. years 1960-1972) that were larger (but with high uncertainty) than the historic SSB predictions from the VPA. The VPA model also predicted a higher aveF in the early 1990's, at approximately 1.1, with the SSM prediction at approximately 0.75 for the same period. From 2005-2009, the estimates for aveF from the SSM were often twice as large as the estimates from the VPA. The retrospective patterns for the SSM were reduced for SSB and greatly reduced for aveF (Fig. 3.8) compared to the VPA.

Table 3.3: Model estimates of variance parameters and some population parameters with percent coefficient of variation. (M1 is base model, M2 is base model with natural mortality=0.20, M3 is base model with natural mortality increased by 0.20, M4 is base model with upper catch bounds set to half the base model upper bounds, M5 is base model with catchability freely estimated).

	SSM		**M2**		**M3**		**M4**		**M5**	
	nll = 4516 AIC=9192 BIC=9595		nll = 4534 AIC=9228 BIC=9631		nll = 4489 AIC=9137 BIC=9541		nll = 4546 AIC=9253 BIC=9656		nll = 4479 AIC=9148 BIC=9627	
	Est	**CV**	**Est**	**CV**	**Est**	**CV**	**Est**	**CV**	**Est**	**CV**
σ_{F5}	1.09	40	1.37	39	0.97	34	1.09	38	0.43	62
σ_{F6^+}	0.81	22	0.86	20	0.94	24	0.83	22	0.52	39
$\sigma_p e$	0.21	6	0.22	6	0.20	6	0.21	6	0.20	6
σ_R	0.70	14	0.98	14	0.65	13	0.67	14	0.64	13
cv_{Fall1}	0.69	20	0.68	18	0.71	20	0.69	20	0.71	19
$cv_{Fall2-5}$	0.33	10	0.33	10	0.34	10	0.33	10	0.34	10
$cv_{Fall6-9}$	0.25	10	0.26	10	0.25	10	0.25	10	0.26	10
$cv_{Fall10-15}$	0.43	9	0.44	9	0.42	9	0.43	9	0.42	9
cv_{Span1}	2.12	47	2.05	46	2.18	48	2.12	47	2.09	46
$cv_{Span2-5}$	1.05	22	1.03	21	1.04	21	1.04	21	1.03	21
$cv_{Span6-9}$	0.55	14	0.54	14	0.54	14	0.54	14	0.53	14
$cv_{Span10-15}$	0.47	12	0.45	11	0.46	12	0.46	12	0.46	12
cv_{Spr1}	1.60	33	1.55	32	1.48	31	1.61	33	1.46	30
cv_{Spr2}	0.82	21	0.80	20	0.75	19	0.82	21	0.75	19
cv_{Spr3-5}	0.47	15	0.45	15	0.44	14	0.48	15	0.44	14
cv_{Spr6-9}	0.29	12	0.28	12	0.27	11	0.29	12	0.27	11
$cv_{Spr10-15}$	0.38	11	0.36	11	0.35	10	0.38	11	0.35	11
ϕ_{FA6^+}	0.99	1	0.99	1	0.99	1	0.99	1	1.00	2
ϕ_{FY5}	0.95	5	0.96	4	0.91	7	0.94	6	0.74	41
ϕ_{FY6^+}	0.94	3	0.95	2	0.94	4	0.92	4	0.87	11
ϕ_{FR}	0.33	45	0.66	16	0.31	47	0.30	49	0.28	52
ϕ_{Fall}	0.57	9	0.59	9	0.58	9	0.56	9	0.59	9
ϕ_{Span}	0.72	6	0.71	6	0.71	6	0.71	6	0.72	6
ϕ_{Spr}	0.88	3	0.87	3	0.86	3	0.88	3	0.86	3
ϕ_{CY}	0.25	29	0.28	27	0.24	30	0.24	31	0.43	23
ϕ_{CA}	0.81	4	0.81	4	0.80	4	0.80	4	0.86	3
F_{2017}	0.08	19	0.12	16	0.02	20	0.07	18	0.02	26
SSB_{2017}	13.63	18	9.59	15	48.01	18	13.88	17	49.19	25
F_{VPA}	0.06	-	-	-	-	-	-	-	-	-
SSB_{VPA}	18.24	-	-	-	-	-	-	-	-	-

Figure 3.4: Model predicted process errors at age for American plaice in NAFO Divisions 3LNO.



The self-simulation study lower 10% and upper 90% intervals for both SSB and aveF covered zero (Fig. 3.9), indicating that the simulated samples produced estimates that were similar to the SSM estimates. In the earliest years (1960-1972), the range of relative differences for aveF was mostly positive, with the converse seen for SSB. Results from the sensitivity tests (Table 3.3) showed that the SSM had the second lowest BIC overall with M3 (natural mortality rates increased by 0.20) having the lowest BIC. The jittered-start test did not converge for 6% of the simulations, with 77% of the converged models producing negative log-likelihoods that were identical to the original formulation.

Figure 3.5: Retrospective estimates from 2008-2017 for total abundance, spawning stock biomass, average fishing mortality rates (ages 9-14) and recruitment for years 1960-2017 for American plaice in NAFO Divisions 3LNO.

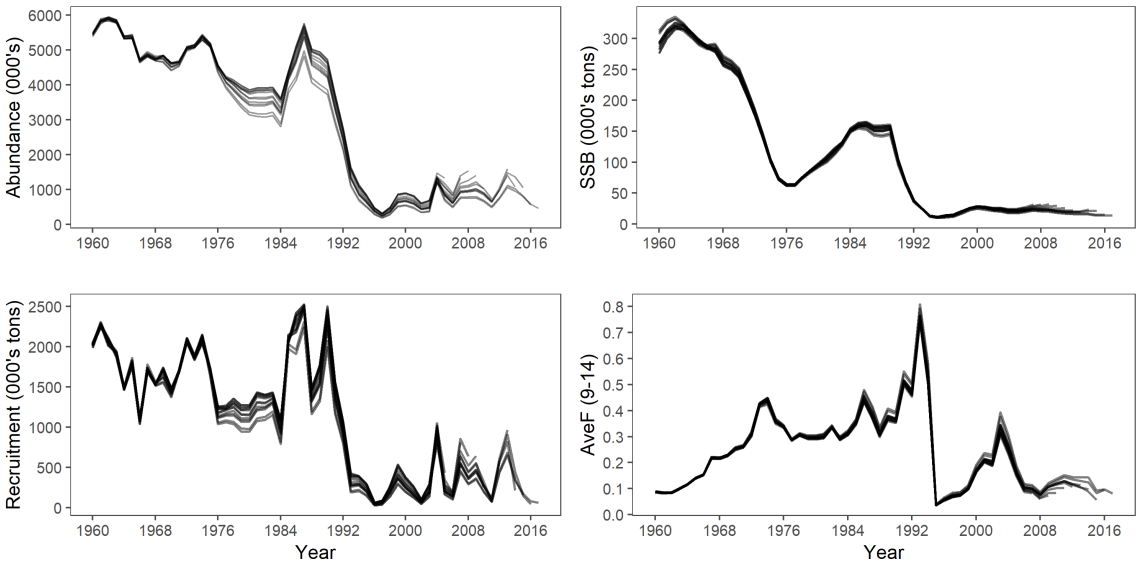


Figure 3.6: Model (SSM) and VPA estimated spawning stock biomass for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent 95% confidence intervals.

Figure 3.7: Model (SSM) and VPA estimated average fishing mortality rates (ages 9-14) for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent 95% confidence intervals.

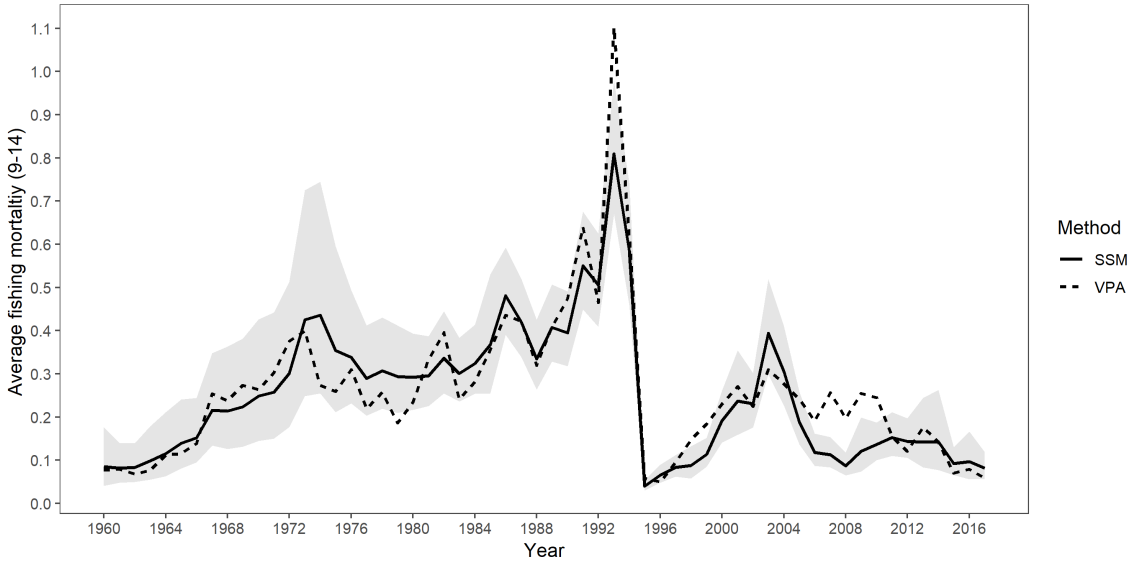


Figure 3.8: Model (SSM) and VPA retrospective estimates from 2013-2017 of average fishing mortality rates (ages 9-14) and of spawning stock biomass for years 2014-2017 for American plaice in NAFO Divisions 3LNO.

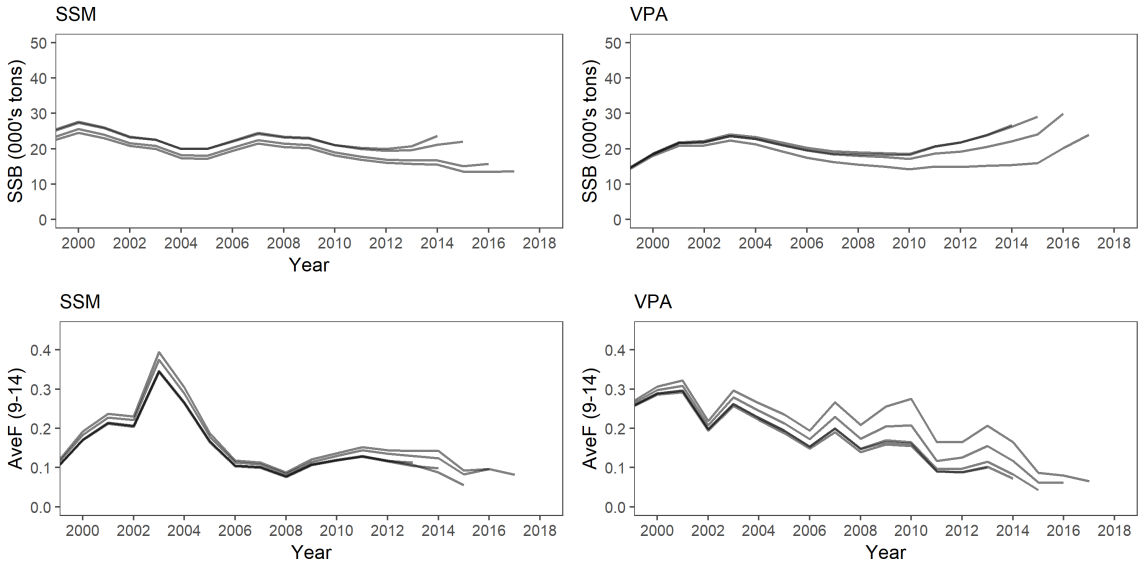
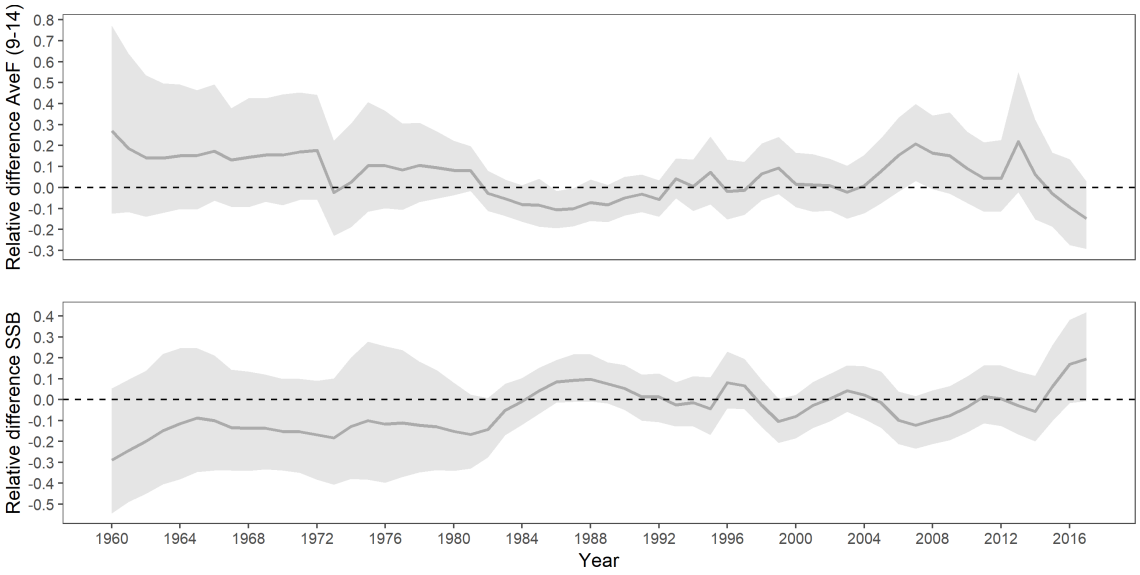


Figure 3.9: Relative difference from model and self-simulated sample for spawning stock biomass and average fishing mortality rates (ages 9-14). The solid grey line in the median of the estimates and shaded grey regions represent the lower 10% and upper 90% bounds.



3.4 Discussion

Overall, our state-space model (SSM) that accounted for uncertainties in the landings data and allowed for process errors fit the data well, with no obvious patterns in the survey and continuation ratio logit residual plots. The retrospective patterns were reduced for spawning stock biomass (SSB) and greatly reduced for average fishing mortality for ages 9-14 (aveF) compared to the most recent stock assessment model (VPA).

The sensitivity runs that reduced M to 0.20 for all ages and years (M2), halved the catch upper bounds (M4), and freely estimated the survey catchabilities (q, M5) all had larger BICs than the SSM. The run that increased the base assumption of M by 0.20 (M3) had a slightly lower BIC than the SSM and this may suggest that the values we used for M's may be too low. Previous research found evidence that M's

during 1989 to 1996 (Morgan & Brodie 2001) had increased to 0.53 and the current VPA model and our SSM include this increase. However, since the closure of the commercial fishery, estimates of total mortality rates have remained high for some periods (e.g. Fig. 3.1 for years 2000-2006), and this may suggest that M is higher than 0.20 in recent years. Preliminary work also suggests that M has increased since the closure of the commercial fishery (COSEWIC 2009, Morgan et al. 2011). Thus, research that improves our understanding of M for this species should be of high priority.

The SSM retrospective analyses indicated very little systematic patterns, which is a key improvement compared to the VPA model. Including process error in the population dynamics model helped account for underlying time-varying population processes that were not accounted for in the VPA, thereby reducing retrospective patterns. There is still evidence of slight retrospective patterns, and this may be caused by underlying spatial or time-varying process that are mis-specified in the observation model since process errors can only account for mis-specifications in the process equations.

The estimate for survey catchability q is defined as the value required to scale swept-area abundance to the population abundance (see, e.g., Dickson 1993, Fraser et al. 2007). An estimate of q less than one implies that fewer fish are caught than occupied the area of the trawl, and a value greater than one implies that more fish are caught than occupied the area. Herding behavior of flatfishes in the presence of survey trawls underestimates the width used in area swept calculations and can result in q estimates that are greater than one (Bryan et al. 2014). Therefore, larger q estimates are not unrealistic for American plaice; however, the q estimates from the SSM are very large, with the maximum estimated at 9.1, which is smaller than

the maximum q estimate from the VPA at 13.6[1] (Table 26, Wheeland et al. 2018). Thus, additional research is required to better understand why the survey swept area abundance estimates are so much higher than the stock assessment model estimates.

A difference to note between the SSM and the VPA is that the SSM assumes that the survey indices are from a normal distribution with a constant coefficient of variation whereas the VPA assumes that the log of the survey indices are from a lognormal distribution. The lognormal distribution does not allow for zeros in the survey data, however this assumption may not be appropriate when there are many zeros in the data or when zeros are "true" zeros (i.e. no fish available to be caught). The assumption of normality with a constant coefficient of variation avoids the problem of dropping zeros altogether. Although the methods are theoretically similar, future research is needed to compare the performance of the two.

Our new SSM is an improvement to the current stock assessment model that is used to inform the management of American plaice on the Grand Bank of Newfoundland as it allows for errors in the landings data and reduces the retrospective patterns. Our results also suggest that the current values used for natural mortality rates may be too low as our diagnostic model fitting found the best model fit when M was increased by 0.2. Specification of M is an important problem in fish stock assessment (Brodziak et al. 2011). We suggest that M assumptions should be rigourously examined, e.g. via M profile plots, to provide a better understanding of model behavior for various assumptions about M. Not only will this allow for a deeper understanding of the role of M in the stock assessment model, but it can provide motivation for research (e.g. Miller & Hyun 2018) into more realistic values of M for future stock assessment models.

[1]Note that the survey index from the NAFO assessment is in millions and the catch is in thousands; to get the qs on the same scale as the SSM we multiplied the NAFO q estimate by 1000

3.5 Acknowledgements

Research funding was provided by the Ocean Frontier Institute, through an award from the Canada First Research Excellence Fund. Research funding to NC was also provided by the Ocean Choice International Industry Research Chair program at the Marine Institute of Memorial University of Newfoundland. Funding to AP was also provided by a Natural Sciences and Engineering Research Council of Canada Master's Graduate Scholarship. Many thanks are also extended to Dr. Anders Nielsen, Danish Technical University, for advice on more computationally efficient ways to implement our model in TMB.

Bibliography

Aeberhard, W. H., Mills Flemming, J. & Nielsen, A. (2018), 'Review of state-space models for fisheries science', Annual Review of Statistics and Its Application **5**, 215–235.

Agresti, A. (2003), Categorical data analysis, Vol. 482, John Wiley & Sons.

Albertsen, C. M., Nielsen, A. & Thygesen, U. H. (2016), 'Choosing the observational likelihood in state-space stock assessment models', Canadian Journal of Fisheries and Aquatic Sciences **74**(5), 779–789.

Berg, C. W. & Kristensen, K. (2012), 'Spatial age-length key modelling using continuation ratio logits', Fisheries Research **129**, 119–126.

Brodziak, J. K. T., Ianelli, J. N., Lorenzen, K. & Methot, R. D. (2011), 'Estimating natural mortality in stock assessment applications august 11-13, 2009, alaska fisheries science center, seattle, wa'.

Bryan, D. R., Bosley, K. L., Hicks, A. C., Haltuch, M. A. & Wakefield, W. W. (2014), 'Quantitative video analysis of flatfish herding behavior and impact on effective area swept of a survey trawl', Fisheries Research **154**, 120–126.

Cadigan, N. G. (2015), 'A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates', Canadian Journal of Fisheries and Aquatic Sciences **73**(2), 296–308.

COSEWIC (2009), 'Cosewic assessment and status report on the american plaice hippoglossoides platessoides, maritime population, newfoundland and labrador population and arctic population, in canada.', Committee on the Status of Endangered Wildlife in Canada pp. 1–74.

Dickson, W. (1993), 'Estimation of the capture efficiency of trawl gear. i: Development of a theoretical model', Fisheries Research **16**(3), 239–253.

Dwyer, R. R., Ings, D., D. Power, M. Morgan, B. B. & Healy, P. (2016), 'An assessment of american plaice in nafo div. 3lno', NAFO SCS Doc **16**, 030.

Fournier, D. & Archibald, C. P. (1982), 'A general theory for analyzing catch at age data', Canadian Journal of Fisheries and Aquatic Sciences **39**(8), 1195–1207.

Fraser, H. M., Greenstreet, S. P. & Piet, G. J. (2007), 'Taking account of catchability in groundfish survey trawls: implications for estimating demersal fish biomass', ICES Journal of Marine Science **64**(9), 1800–1819.

González-Troncoso, D. M., Román-Marcote, E. & Paz, X. (2017), 'Results for greenland halibut, american plaice and atlantic cod of the spanish survey in nafo div. 3no for the period 1997-2016'.

Hilborn, R. & Walters, C. J. (2013), Quantitative fisheries stock assessment: choice, dynamics and uncertainty, Springer Science & Business Media.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. & Bell, B. (2015), 'Tmb: automatic differentiation and laplace approximation', arXiv preprint arXiv:1509.00660 .

Legault, C. M. (2009), 'Report of the retrospective working group, january 14-16, 2008, woods hole, massachusetts'.

Maunder, M. N. & Punt, A. E. (2013), 'A review of integrated analysis in fisheries stock assessment', Fisheries Research **142**, 61–74.

Megrey, B. A. (1989), Review and comparison of age-structured stock assessment models, in 'Amer. Fish. Symp.', Vol. 6, pp. 8–48.

Miller, T. J. & Hyun, S.-Y. (2018), 'Evaluating evidence for alternative natural mortality and process error assumptions using a state-space, age-structured assessment model', Canadian Journal of Fisheries and Aquatic Sciences **75**(5), 691–703.

Mohn, R. (1999), 'The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data', ICES Journal of Marine Science **56**(4), 473–488.

Morgan, M. & Brodie, W. (2001), 'An exploration of virtual population analyses for divisions 3lno american plaice', NAFO, NAFO SCR Doc **1**(4).

Morgan, M. J., Bailey, J., Healey, B., Maddock Parsons, D. & Rideout, R. (2011), Recovery potential assessment of American Plaice (Hippoglossoides platessoides) in Newfoundland and Labrador, Fisheries and Oceans Canada, Science.

Nielsen, A. & Berg, C. W. (2014), 'Estimation of time-varying selectivity in stock assessments using state-space models', Fisheries Research **158**, 96–101.

Pitt, T. (1972), 'Nominal catches of american plaice in divisions 3l and 3n for the years 1960-1970', International Commission for the Norwest Atlantic Fisheries **72**(90).

Shelton, P. A. & Morgan, M. J. (2005), 'Is by-catch mortality preventing the rebuilding of cod (gadus morhua) and american plaice (hippoglossoides platessoides) stocks on the grand bank', J. Northw. Atl. Fish. Sci **36**, 1–17.

Skaug, H. J. & Fournier, D. A. (2006), 'Automatic approximation of the marginal likelihood in non-gaussian hierarchical models', Computational Statistics & Data Analysis **51**(2), 699–709.

Szuwalski, C. S., Ianelli, J. N. & Punt, A. E. (2017), 'Reducing retrospective patterns in stock assessment and impacts on management performance', ICES Journal of Marine Science **75**(2), 596–609.

Wheeland, L., Dwyer, K., Morgan, M., Rideout, R. & Rogers, B. (2018), 'Assessment of american plaice in div. 3lno', Évaluation .

Chapter 4

Summary and Papers

The overarching goal of this thesis was to improve age-based stock assessment models with an application to American plaice on the Grand Bank of Newfoundland. In Chapter 1, I reviewed age-based stock assessment models from their first appearance in the fisheries literature up to today and touched on current data-related issues both generally and with a specific focus on American plaice.

In Chapter 2, I collated the methods found in the fisheries science and statistical literature that aim to account for the length-stratified age sampling design when estimating growth parameters. Nine methods were studied and a thorough simulation study was conducted to determine the best performing method. My results showed that the most commonly used methods in fisheries science (i.e. bias-corrected and weighted bias-corrected) had poor accuracy and the empirical proportion approach was optimal. This approach was applied to Grand Bank American plaice.

In Chapter 3, I developed an age-based state-space stock assessment model (SSM) for Grand Bank American plaice. This model allowed for errors in the underlying population processes and in the catch-at-age data. This is a more realistic formulation than the current assessment model (VPA) as there is considerable uncertainty about the landings data for this stock and the catch-at-age data are derived in part from landings estimates. The SSM provided similar estimates of population productivity as the VPA model but with improvements in the retrospective plots. The retrospective

patterns for spawning stock biomass were reduced and greatly reduced for average fishing mortality rates in comparison to the VPA.

This research not only improved our understanding of Grand Bank American plaice population dynamics and growth but is also a step forward in fitting more realistic integrated age-based stock assessment models. In future work, growth models that account for the length-stratified age sampling design can be combined with length-based indices and age-length measurements to fit integrated stock assessment models with separate likelihood functions. These future models that allow for variability in all data sources and avoid subjective data-weighting will have potentially far reaching impacts as they can provide a more reliable assessment from which fisheries managers can make decisions (Punt et al. 2020). Secondly, our state-space stock assessment model for American plaice that fit a separate likelihood for the landings data and catch proportions can be a useful template for researchers interested in fitting integrated age-based state-space models to similar stocks. Allowing for important variability in the data can greatly reduced the retrospective patterns, and this has the potential to increase the confidence in the assessment output that is being provided to fisheries managers.

4.1 Paper 1

NRC Research Press

ARTICLE

Estimation of growth parameters based on length-stratified age samples

Andrea M.J. Perreault, Nan Zheng, and Noel G. Cadigan

Abstract: Response-selective stratified sampling (RSSS) has been well studied in the statistical literature; however, the application of the resulting statistical theories and methods to a specific case of RSSS in fisheries studies, namely length-stratified age sampling (LSAS), is inadequate. We review nine estimation approaches for RSSS found in the statistical and fisheries science literature in terms of three sampling components: the first phase length composition sample, the second phase age composition sample, and the sampling scheme. We compare the performance in terms of RRMSE (relative root mean squared error) for von Bertalanffy (vonB) growth model parameter estimation using an extensive simulation study. We further demonstrate methods by applying the two best-performing and the most popular methods to estimate the vonB model parameters for American plaice (*Hippoglossoides platessoides*) in NAFO Divisions 3LNO. Our simulations demonstrated that mis-specifying one or more of the three sampling components increases the RRMSEs, and this effect is magnified when the age distribution is incorrectly specified. The optimal approach for data based on LSAS is the empirical proportion approach, and we recommend this method for growth parameter estimation based on LSAS data.

Résumé : L'échantillonnage stratifié à sélectivité basée sur la réponse (ESSR) est un sujet bien étudié dans la documentation statistique, mais l'application des théories et méthodes statistiques en découlant à un cas précis d'ESSR dans les études sur les pêches, à savoir l'échantillonnage d'âges stratifié selon la longueur (EASL), est inadéquate. Nous passons en revue neuf approches d'estimation pour l'ESSR relevées dans la documentation statistique et sur les sciences halieutiques au vu de trois composants de l'échantillonnage, soit l'échantillon de la composition des longueurs de la première phase, l'échantillon de la composition des âges de la deuxième phase et le plan d'échantillonnage. Nous en comparons la performance à la lumière de l'erreur quadratique moyenne relative (EQMR) pour l'estimation des paramètres du modèle de croissance de von Bertalanffy (vonB) en utilisant une vaste étude de simulation. Nous faisons en outre la démonstration de méthodes en appliquant les deux méthodes les plus performantes et les méthodes les plus populaires pour estimer les paramètres du modèle vonB pour la plie canadienne (*Hippoglossoides platessoides*) dans les divisions 3LNO de l'OPANO. Nos simulations démontrent que la définition erronée d'au moins un des trois composants de l'échantillonnage accroît les EQMR, et cet effet est amplifié si la répartition des âges est définie incorrectement. L'approche optimale pour des données basées sur l'EASL est l'approche des proportions empiriques, et nous recommandons cette méthode pour l'estimation des paramètres de croissance basée sur des données d'EASL. [Traduit par la Rédaction]

Introduction

Modeling and estimation of the relationship between fish length and age (i.e., growth) is an important component of fisheries science. Growth models are used in many ways, including the derivation of life history invariants (Charnov 1993; Jensen 1996), estimation of important population parameters such as selectivity and mortality (Hoggarth 2006; Hilborn and Walters 2013), and the classification of functional groups in ecosystems models (e.g., Shackell et al. 2010). Growth models are especially important in age-based fish stock assessments, where the estimates from these models are used in conjunction with a length–weight relationship in biomass calculations (Quinn and Deriso 1999), to convert length-based to age-based selectivity, and to estimate length compositions (Francis 2016). Stock assessment scientists must decide whether growth will be estimated inside the stock assessment model via integrated stock assessment models (Methot and Wetzel 2013; Maunder and Piner 2015) or outside the stock assessment model, where the fixed growth parameter estimates are used as data inputs for the assessment. Alternatively, empirical lengths and weights at age derived from individual fish measurements can also be applied instead of growth model estimates; however, this method requires high-quality data across all assessment model years, which are often not available. Thus, reliable methods to estimate growth model parameters are vital, as unreliable management decisions and conclusions can be drawn about the health of a fish stock if the growth estimates are incorrect (Smith et al. 1993).

The most widely used growth model in fisheries science is the von Bertalanffy model (vonB; von Bertalanffy 1938), which assumes that the growth rate decreases linearly as size increases. The vonB parameters are estimated from samples of age and length measurements. Usually, many more fish are measured for length than age because determining age is often time-consuming and expensive (Doubleday 1981), whereas lengths are much simpler to measure. A subsample of fish is selected for ageing. A popular sampling design for this is length-stratified age sampling (LSAS), which is a two-phase stratified sampling design. In the first sampling phase, length is measured for the entire random sample of fish, and the fish are classified into length bins, or strata

Received 8 April 2019. Accepted 23 July 2019.

A.M.J. Perreault, N. Zheng, and N.G. Cadigan. Centre for Fisheries Ecosystems Research, Marine Institute of Memorial University of Newfoundland, P.O. Box 4920, St. John's, NL A1C 5R3, Canada.

Corresponding author: Andrea M.J. Perreault (email: andrea.perreault@mi.mun.ca).

(e.g., 1 cm, 5 cm). In the second sampling phase, a small and prespecified number of fish are randomly selected to be aged from each length stratum. The rationale for this design is to reduce the number of ages to measure but maintain the representativeness of the data by ensuring enough age measurements in all the length intervals along the growth curve. This is an example of response-selective stratified sampling (RSSS), since length is the response and age is the descriptive covariate. There are many well-studied factors that can bias vonB parameter estimates, including selectivity (e.g., Troynikov 1999; Taylor et al. 2005; Schueller et al. 2014), age measurement error (Cope and Punt 2007), and between-individual variability (see Sainsbury 1980; Kimura 2008; Vincenzi et al. 2016); however, bias from LSAS has not been sufficiently studied.

Although RSSS has been intensively studied, especially during the last two or three decades in the statistical literature (see, e.g., Neyman 1938; Hsieh et al. 1985; Breslow et al. 2009; Scott and Wild 2011), the application of the resulting statistical theories and methods to LSAS in fisheries science is inadequate. For example, popular and easy-to-use approaches such as the weighted likelihood and calibrated weighted likelihood methods (detailed later; see Kalbfleisch and Lawless 1988; Saegusa and Wellner 2013) in RSSS are seldom applied in fisheries studies, and the few studies to address LSAS have various theoretical or empirical issues to resolve. In the latter regard, Candy et al. (2007) accounted for LSAS with an approach for a different kind of RSSS, the variable probability sampling design, which still needs some theoretical and (or) simulation validation. Another approach estimates mean length at each age, adjusted for LSAS through weights, called the bias-corrected mean length-at-age (BC) estimates (e.g., Bettoli and Miranda 2001). The BC estimates are treated as observed data and used to estimate growth model parameters using the nonlinear least squares method (nls; e.g., see Echave et al. 2012; Brattey et al. 2018); however, the effect of using the BC estimates and not the individual observations on growth parameter estimates still requires further investigation. In addition, it is not possible to estimate between-individual variation in size-at-age with the BC approach, which is an important limitation of the method for some stock assessment models that require estimates of the stock distribution of size-at-age. Piner et al. (2016) and Lee et al. (2017) used length, instead of age, as the covariate by conditioning on length to circumvent the complexity with RSSS, since given length, the subsampling for ageing is random in LSAS. However, in this paper we show this conditional method is both theoretically problematic and practically unreliable. On the other hand, the theories and approaches proposed from statistical studies need some further development because they do not fully account for the specific complexities of fisheries surveys. For example, in LSAS, the number of length strata (100–200) is far more than that typical in the statistical literature on RSSS (~10), and in LSAS, there are always some strata that are theoretically probable but practically empty, which is mostly neglected in statistical studies.

Zheng and Cadigan (2019) addressed these additional complexities in fisheries surveys by proposing the full-data likelihood function for LSAS, and they also improved the method of Candy et al. (2007), which they referred to as the empirical proportion likelihood approach. Their simulation study indicated that these two new approaches perform as well as the standard full information likelihood approach and better than the other existing approaches in RSSS. An important goal of this paper is to compare the new approaches of Zheng and Cadigan (2019) with the existing methods for LSAS in the fisheries literature to determine one or two optimal LSAS vonB parameter estimation procedures. Zheng and Cadigan (2019) assumed that an appropriate model for the age distribution of the population is known, which is usually not the case for fish stocks, as seen from our simulation study. Another goal of this paper is to address this deficiency. Zheng and Cadigan (2019) concluded that for efficient and consistent parameter estimation when data are collected using LSAS, information from the following three components must be included: the partially observed (i.e., only length) first phase data, the fully observed (i.e., length and age) second phase data, and the sampling scheme. In this paper, we guide our estimating method comparison in light of these three components.

In the rest of this paper, we present the estimation performance of the vonB parameter estimates using nine estimation methods from the fisheries science and statistical literature. As a practical application, the two best-performing methods were used to fit the vonB growth model for American plaice (*Hippoglossoides platessoides*) in Northwest Atlantic Fisheries Organization (NAFO) Divisions 3LNO.

Materials and methods

Suppose we have N randomly selected fish with lengths and ages (l_i, a_i) for $i = 1, 2, ..., N$, generated from the joint distribution $f(l, a \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of all unknown parameters. The individuals are divided into K mutually exclusive and exhaustive length strata, $S_1, S_2, ..., S_K$. We measure all N fish for length in the first phase sample, but only a subsample of size n is measured for age in the second phase. Thus, the full age and length information (l_i, a_i) is only collected for a subset of the N fish. The probability that a fish with length l falls into the kth stratum (i.e., $l \in S_k$, $k = 1, 2, \ldots, K$) is given by

$$(1) \qquad Q_k(\boldsymbol{\theta}) = P[(l, a) \in S_k \mid \boldsymbol{\theta}] = \int_{l \in S_k} \int_a f(l, a \mid \boldsymbol{\theta}) \mathrm{d}a \, \mathrm{d}l$$

We denote the number of fish in a stratum by N_k, where $N = \sum_{k=1}^{K} N_k$. In each stratum, we specify a maximum number of fish to be sampled for age, m_k. The true sample size n_k for this stratum is

$$(2) \qquad n_k = \begin{cases} N_k, & \text{if } N_k < m_k \\ m_k, & \text{if } N_k \geq m_k \end{cases}$$

which is a random number. This is called LSAS. In fisheries surveys, it is rare to sample the target m_k fish in strata for smaller and larger lengths. Fisheries survey gears may allow most smaller fish to escape, and it is uncommon to catch many fish of the oldest ages. As a result, the distribution of fish length is often highly right-skewed, and in this case the n_k values are random variables.

For LSAS (eq. 2), the density function for the second phase length-at-age data is not $f(l, a \mid \boldsymbol{\theta})$ anymore, but given by (see Appendix A for derivation)

$$(3) \qquad f(l, a \mid \mathrm{LSAS}, \boldsymbol{\theta}) = \frac{f(l, a \mid \boldsymbol{\theta})}{Q_k(\boldsymbol{\theta})} Q_k(\mathrm{LSAS}, \boldsymbol{\theta})$$

where $Q_k(\mathrm{LSAS}, \boldsymbol{\theta})$ is the probability for a second phase individual to be in length stratum S_k. Intuitively, $Q_k(\mathrm{LSAS}, \boldsymbol{\theta})$ is related to the distribution of the random variable n_k. For instance, given n_k for each stratum, a second phase individual is twice as likely to be in a stratum with $n_k = 2$ than in a stratum with $n_k = 1$; if n_k was to change for a stratum, then $Q_k(\mathrm{LSAS}, \boldsymbol{\theta})$ for the second phase stratum would also change. Zheng and Cadigan (2019) derived the exact formula for $Q_k(\mathrm{LSAS}, \boldsymbol{\theta})$, which not only validates the above observation about the relationship between $Q_k(\mathrm{LSAS}, \boldsymbol{\theta})$ and the distribution of n_k, but also indicates that $Q_k(\mathrm{LSAS}, \boldsymbol{\theta})$ is indeed a function of the target parameters $\boldsymbol{\theta}$ and hence should not be neglected for $\boldsymbol{\theta}$ estimation. Their simulation study further confirmed that approaches neglecting $Q_k(\mathrm{LSAS}, \boldsymbol{\theta})$ give fairly poor $\boldsymbol{\theta}$ estimates.

Table 1. Log-likelihood functions used to estimate von Bertalanffy (vonB) population parameters.

Method	Likelihood	$P(a)$	First phase	Sampling scheme
Random	$\sum_l \sum_a n_{l,a} \cdot \log[P(l \mid a)P(a)]$	Crl	No	No
Piner	$\sum_l \sum_a n_{l,a} \cdot \log\left[\frac{P(l \mid a)\hat{P}(a)}{P(l)}\right]$	Equil.	No	Incorrect; neglects $\boldsymbol{\theta}$ in $P(l)$
PF	$\sum_l \sum_a n_{l,a} \cdot \log\left[\frac{P(l \mid a)P(a)}{Q_k}Q_k(\text{LSAS})\right]$	Crl	No	Yes
HT	$\sum_l \sum_a n_{l,a} \cdot \pi_{k(l)} \cdot \log[P(l \mid a)P(a)]$	Crl	N_k	Yes, through weights
CW	$\sum_l \sum_a n_{l,a} \cdot w_i \cdot \log[P(l \mid a)P(a)]$	Crl	N_k, l	Yes, through weights
Candy	$\sum_l \sum_a n_{l,a} \cdot \log\left[\frac{P(l \mid a)}{\sum_k [n_{k(l)}/N_{k(l)}]Q'_k(a)}\right]$	NA	N_k	Yes, ignores empty strata
EP	$\sum_l \sum_a n_{l,a} \cdot \log\left\{\frac{P(l \mid a)P(a)}{\sum_{k'=1}^{K_{\text{obs}}} [n_{k(l)'}/N_{k(l)'}]Q_{k'} + \sum_{k'=K_{\text{obs}}+1}^{K_{\text{total}}} Q_{k'}}\right\}$	Crl	N_k	Yes, through EP density

Note: PF, partial-full; HT, Horvitz–Thompson; CW, calibrated-weighted; EP, empirical proportion; Crl, continuation ratio logit transformation; and Equil., equilbrium approximation in eq. 8. Bias-corrected mean length-at age (BC) and weighted BC (WBC) methods are not listed, as they are not likelihood approaches. NA, not applicable.

In the special case where the probability that $N_k < m_k$ is negligible for each stratum, then $n_k = m_k$ with probability 1 for each stratum. Namely, the n_k values are fixed numbers instead of random variables, and $Q_k(\text{LSAS}, \boldsymbol{\theta}) = m_k / \sum m_k$. This probability does not involve target parameters $\boldsymbol{\theta}$ and hence can be neglected for the purpose of $\boldsymbol{\theta}$ estimation. Some approaches (e.g., Hausman and Wise 1982) basically assume that the n_k values are always equal to m_k values and hence neglect the $Q_k(\text{LSAS}, \boldsymbol{\theta})$ term in eq. 3.

From the point of view of the three essential components for consistent and efficient parameter estimation, namely the N first phase length data, the n second phase length-at-age data, and the sampling design, neglecting $Q_k(\text{LSAS}, \boldsymbol{\theta})$ is an incorrect incorporation of the sampling design and hence leads to poorer estimation results. Zheng and Cadigan (2019) suggested that all the deficiencies in RSSS estimation approaches to LSAS are due to insufficient or incorrect incorporation of one or more of these components. For the methods that we investigate in this paper, the first phase information can be incorporated via the number of fish in each length bin (N_k) or the length of each fish. Incorporation of the sampling scheme indicates whether some measure has been taken to eliminate or alleviate the bias in the data introduced by LSAS. In the remainder of this paper, we will categorize and analyze the existing LSAS estimation approaches in light of whether they incorporate these three components correctly and sufficiently.

Model and likelihoods

Length-at-age is commonly described by the vonB growth model (e.g., Francis 2016; Quist et al. 2012), as it gives a good approximation of growth for many fish species (e.g., Chen et al. 1992; Quinn and Deriso 1999). The vonB model can be written as

$$l(a) = L_\infty[1 - e^{-k(a-a_0)}] + \epsilon \quad (4)$$

where $l(a)$ is the length of an age a fish, L_∞ is the theoretical length at which the growth rate stops, k is the growth coefficient, and a_0 is the theoretical age at which length is zero. We assume the error term is given by $\epsilon \sim N(0, \sigma_a^2)$, where $\sigma_a = \mu_a \text{CV}$, CV is the coefficient of variation, and $\mu_a = L_\infty[1 - e^{-k(a-a_0)}]$. We assume this error term jointly represents the between-individual variation (or process error) resulting from individual growth parameter variability, plus the measurement error in length, as in Piner et al. (2016). Note that it is straightforward to apply the estimation methods we investigate to other growth models (e.g., the Gompertz model).

Equation 4 in fact states that the distribution of length l conditional on a, $f(l \mid a) \sim N(\mu_a, \sigma_a^2)$. In practice, real length and age data are discrete integers, so the corresponding probability mass function (pmf) can be approximated as

$$P(l \mid a) = \frac{N(l, \mu_a, \sigma_a^2)}{\sum_l N(l, \mu_a, \sigma_a^2)}$$

To fit five of the methods described below (see Table 1), the population distribution of age, $P(a)$, is required. In practice, the age distribution of a fish stock may be complicated and will change substantially over time because of the highly variable nature of fish recruitment. Simple models will usually not be reliable. In preliminary simulations (see section below on Simulation study), we tested many common probability distributions for $P(a)$, including the Poisson, Gamma, and G-Normal mixture distributions as in Dey et al. (2019), but none of them performed well, as the randomly generated age distributions for our simulations (described in section on Simulation study) varied from year to year. Hence, we decided to estimate the pmf $P(a)$ at all ages directly. We used the continuation ratio-logit transformation (e.g., Cadigan 2016; Berg and Kristensen 2012; Agresti 2003), which maps $P(a)$ for $a = 1, \ldots, A_{\max}$ with the constraints $P(a) \geq 0$ and $\sum P(a) = 1$, into $\lambda(a)$ for $a = 1, \ldots, A_{\max} - 1$ with no constraints, which are much easier to estimate. Thus,

$$P(a) = \begin{cases} \dfrac{e^{\lambda(a)}}{\prod_{i=1}^{a}\{1 + e^{\lambda(i)}\}} & a = 1, \ldots, A_{\max} \\ \dfrac{1}{\prod_{i=1}^{A_{\max}-1}\{1 + e^{\lambda(i)}\}} & a = A_{\max} \end{cases} \quad (5)$$

where $A_{\max}$ is the maximum age group for age groups $a = 1, \ldots, A_{\max}$, and $\lambda(a) \in (-\infty, \infty)$ for $a = 1, \ldots, A_{\max} - 1$. The inverse transformation of eq. 5 is

$$\lambda(a) = \log\left[\frac{P(a)}{P(a+1) + \ldots + P(A_{\max})}\right], \quad a = 1, \ldots, A_{\max} - 1 \quad (6)$$

When implementing the various methods discussed in the following sections, $P(a)$ should be replaced by the functions of $\lambda(a)$ values in eq. 5, and $\lambda(a)$ values are estimated together with the other model parameters.

Random method

Often in practice, because users are either not aware of, or do not understand the significance of, how their age–length data

were collected, the LSAS design is inadvertently ignored and the second phase length and age sample are treated as a random sample from the population when estimating vonB parameters. We used this approach to demonstrate how biased and unreliable growth model parameter estimates can be when this approach is applied to LSAS data. The most common way to implement this method fits the second phase length and age data to eq. 4, and the vonB parameters are estimated using nls with age treated as a fixed covariate. The random likelihood method described in Table 1 will give identical parameter estimates to the nls method when the distribution of age does not depend on the vonB parameters of interest. This method ignores both the first phase information and the sampling scheme, since it treats the second phase sample as a random sample.

Weighted methods

One of the most common approaches for statistical inference in RSSS is by weighing the likelihood function for the second phase data with information from the first phase sample (see Kalbfleisch and Lawless 1988; Breslow et al. 2009; Saegusa and Wellner 2013). These weighted methods are relatively easy to implement and can provide unbiased parameter estimating equations. The commonly used Horvitz–Thompson (HT) weight, $\pi_{k(l)} = N_{k(l)} / n_{k(l)}$, where $k(l)$ represents length l in stratum k, incorporates the first phase information and the sampling scheme through the fraction sampled for ageing ($\pi^{-1}_{k(l)}$; see Table 1 for all likelihood functions). Note that $N_{k(l)}$ contains some of the first phase length information. For example, if fish are measured by centimetre and 1 cm length bins are used for age sampling, then $N_{k(l)}$ values provide all the first phase length information. If 5 cm length bins are used, then the $N_{k(l)}$ values only provide partial first phase length information. The HT approach may not be efficient since the HT estimator may have large standard errors when the sampling units are not proportional to their inclusion probabilities (e.g., Thompson 2012). To adjust for this inefficiency, the calibrated-weighted (CW) method, first proposed by Deville and Särndal (1992), further adjusts the HT weights with first- and (or) higher-order moments of the first phase length data (see Appendix B for details):

$$w_i = (1 - \lambda_c l_i)\frac{N_{k(l)}}{n_{k(l)}} \tag{7}$$

Both the HT and CW methods incorporate some of the first phase information and the sampling scheme through their respective weights.

Conditional-on-length method

Piner et al. (2016) applied a conditional-on-length method based on the conditional probability

$$P(a \mid l) = \frac{P(l \mid a)P(a)}{\sum_a P(l \mid a)P(a)}$$

to estimate the vonB parameters with data collected through LSAS. The rationale is that the conditional distribution of age given length does not depend on the length-stratified sampling design. The corresponding log-likelihood in Piner et al. (2016) is $\Sigma_l \Sigma_a n_{l,a} \cdot \log[P(a \mid l)]$. Note that $P(a, l) = P(a \mid l)\, P(l)$. The corresponding log-likelihood should be $\Sigma_l \Sigma_a n_{l,a} \cdot \log[P(a \mid l)] + \Sigma_l \Sigma_a n_{l,a} \cdot \log[P(l)]$. If $P(l)$ does not include model parameters $\boldsymbol{\theta}$, it can be neglected when applying the maximum likelihood approach. However, $P(l)$ as defined by eq. 4 must involve model parameters, and hence the second term involving $P(l)$ in the log-likelihood should be included. Therefore, the sampling scheme was incorrectly specified in the method of Piner et al. (2016), since it neglected the second term, $P(l)$. In addition, this method did not include any first phase length information. Piner et al. (2016) also used an equilibrium approximation for the distribution of ages

$$\hat{P}(a) = \frac{e^{-Za}}{\sum_a e^{-Za}} \tag{8}$$

where Z is the total mortality rate, defined as the sum of the natural mortality rate (M) and fishing mortality rate (F): $Z = M + F$. This approximation is based on an assumption of constant recruitment of age 1 fish. We note that this is the only method that requires an external estimate of M, as the other methods estimated $P(a)$ within the model (when required; see Table 1) using continuation ratio-logits. This assumption for M is a large simplification, as in reality, natural mortality may not be time- or location-invariant. We investigate whether the equilibrium approximation (eq. 8), which will often not be appropriate to describe the age distribution, is a useful assumption for estimating vonB parameters.

Conditional-on-age method

Candy et al. (2007) accounted for LSAS in vonB parameter estimation by treating the sampling scheme as variable probability sampling (VPS). VPS differs from LSAS in that instead of fixing the stratum subsample size, the subsampling probability of an individual is predetermined. For a discussion on the feasibility of using VPS for a sample collected through LSAS, we refer to Zheng and Cadigan (2019) and Candy et al. (2007). Candy's method differed from the traditional VPS estimation by substituting the observed proportions, $n_{k(l)} / N_{k(l)}$, which are independent of age, for the sampling probabilities. Candy et al. (2007) used the conditional probability given age to avoid the difficulty with the age distribution. The log-likelihood function for this approach is shown in Table 1. The conditional probability of being in the kth stratum is given by

$$Q'_k(a; \boldsymbol{\theta}) = \Pr(l \in S_k \mid a) = \int_{l \in S_k} f(l \mid a)\mathrm{d}l \tag{9}$$

Candy's approach incorporated the first phase information through the observed proportions, $n_{k(l)} / N_{k(l)}$. We consider conditioning on age further in the Discussion.

Improved conditional-on-age and full likelihood methods

Zheng and Cadigan (2019) proposed two likelihood approaches for parameter estimation in LSAS: the full-data likelihood method and the empirical proportion method (EP). Both the full-data and EP methods suggest incorporating the first phase data through the marginal density function of length: $f(l) = \Sigma_a f(l \mid a)P(a)$. However, the first phase density functions can only be correctly evaluated if there exists an appropriate age distribution for integrating a out of $f(l, a \mid \boldsymbol{\theta})$. In a preliminary analysis, we tried to infer the age distribution from the LSAS data; however, the initial results were not promising, and we did not pursue it further. Owing to the lack of an effective method for estimating the age distribution for the simulated populations in this study, we estimated the vonB parameters using only the density functions for the second phase age and length data, which we call the EP method and the partial full-data (PF) method.

The EP approach improved on Candy's method by accounting for the scenarios where no fish are observed in a stratum (i.e., $N_k = 0$; see Table 1) and does not use the conditional-on-age approach (i.e., uses eq. 1). We discuss the drawbacks of using the probability of length conditional on age in the Discussion. The EP joint density function for l and a is

Table 2. Distributions and models used to generate population simulations.

Parameter	Distribution	Mean	SD
Natural mortality (M)	Uniform	0.30	0.11
Fishing mortality (F)	TruncNorm	0.20	0.08
Steepness (h)	Normal	0.75	0.07
Log recruitment deviation (σ_{Rec})	TruncNorm	0.60	0.15
Asymptotic length (L_∞)	Normal	50	8
Growth coefficient error (ϵ)	Normal	0	0.10(1.65/M)
CV length-at-age	TruncNorm	0.10	0.04
Log unfished recruitment [log(R_0)]	Fixed	9.0	—
Length at age 0 (L_0)	Fixed	3	—
Age at length 0 (a_0)	Fixed	−0.07	—

Parameter	Model
Recruitment	$R(y) = \dfrac{\mathrm{SSB}(y-1)}{1-\left(\dfrac{5h-1}{4h}\right)\left[1-\dfrac{\mathrm{SSB}(y-1)}{\mathrm{SSB}_{eq}}\right]}$
Weight	$w[l(a)] = 0.20 \cdot l(a)^3$
Maturity	$m(a) = \left\{1+\mathrm{e}^{\frac{[-\log(19)(a-A_{50})]}{A_{95}-A_{50}}}\right\}^{-1}$
Numbers at age	$N(y, a) = N(y-1, a-1)\mathrm{e}^{-[M+F(y)]}$
Numbers in first year	$N(1, a) = R_0(\mathrm{e}^{-a \cdot M})$
Spawning stock biomass	$\mathrm{SSB}(y) = \sum_{a=1}^{A} w[l(a)] \cdot m(a) \cdot N(y,a)$

Note: Minimum = 0.10 and maximum = 0.50 for the uniform distribution; TruncNorm is the truncated normal distribution with minimum = 0, maximum = ∞ to prevent generating negative values. When a_0 was estimated freely, a_0 was used; otherwise L_0 was used.

$$(10)\qquad f_{\mathrm{EP}}(l, a \mid \mathrm{LSAS};\ \boldsymbol{\theta}) = \frac{[n_{k(l)}/N_{k(l)}]f(l, a \mid \boldsymbol{\theta})}{\sum_{k'=1}^{K_{\mathrm{obs}}}[n_{k'(l)}/N_{k'(l)}]Q_{k'} + \sum_{k'=K_{\mathrm{obs}}+1}^{K_{\mathrm{total}}} Q_{k'}}$$

where 1, …, K_{obs} are the strata with observed data, and K_{obs} + 1, …, K_{total} are the strata without data.

The PF method incorporates the sampling scheme correctly (see Zheng and Cadigan 2019 for full derivation), and the joint density function for l and a with $l \in S_k$ is

$$(11)\qquad f_{\mathrm{PF}}(l, a \mid \mathrm{LSAS};\ \boldsymbol{\theta}) = \frac{f(l, a \mid \boldsymbol{\theta})}{Q_k} \times \frac{\left[\sum_{i=1}^{m_k-1} i \cdot \mathrm{dbin}(i, N, Q_k)\right] + m_k[1 - \mathrm{pbin}(m_k - 1, N, Q_k)]}{\sum_{k'=1}^{K}\left\{\sum\left[\sum_{i=1}^{m_{k'}-1} i \cdot \mathrm{dbin}(i, N, Q_{k'})\right] + m_{k'}[1 - \mathrm{pbin}(m_{k'} - 1, N, Q_{k'})]\right\}}$$

where dbin represents the binomial probability mass function, pbin is the cumulative binomial pmf, and m_k is the maximum sample size for full observation in stratum k. For the EP method, the first phase information was included through the empirical proportions, and the sampling scheme was incorporated through the EP density function. The PF method correctly incorporated the sampling scheme, but only included first phase information via the total sample size N. Note that the total sample size provides very little information about the overall sampling scheme; for example, the difference between N = 1000 in eq. 11 and N = 2000 is negligible.

Mean length-at-age

We also studied fitting the vonB model to mean length-at-age (e.g., Echave et al. 2012) that was adjusted to account for bias due to the LSAS. This approach is commonly used in practice. The bias-corrected mean length-at-age (BC) estimate is

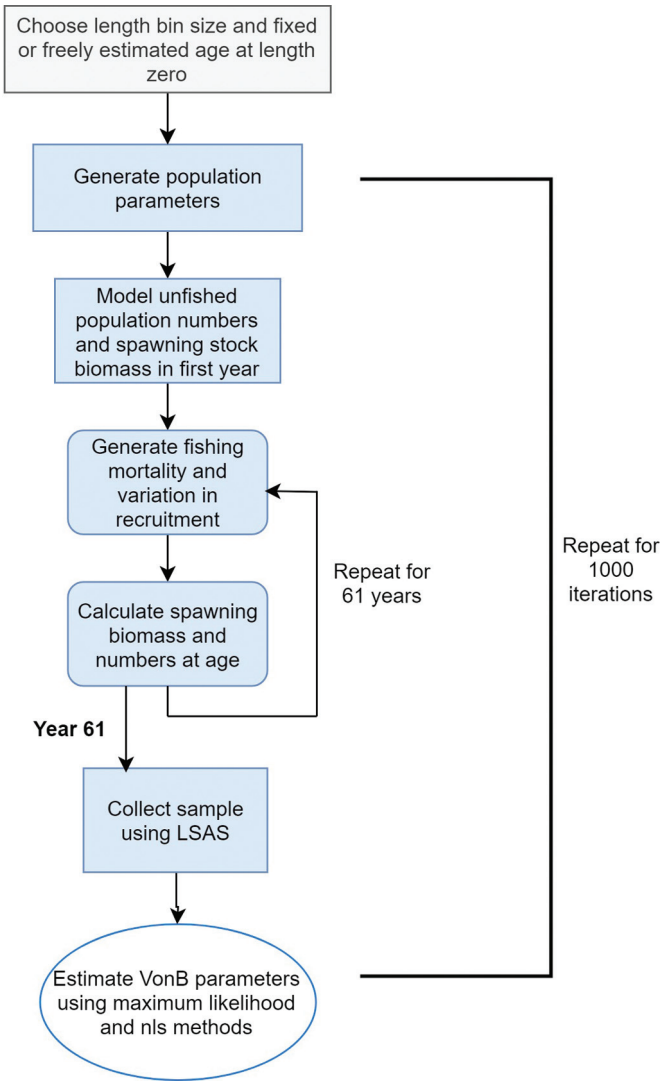
$$(12)\qquad \bar{L}_a = \frac{\sum_k N_k(n_{a,k}/n_k)l_k}{\sum_k N_k(n_{a,k}/n_k)}$$

where l_k is the midpoint of length bin k, $n_{a,k}$ is the number of age a fish in length bin k, and N_k is the total number of fish sampled in length bin k in the first sampling phase. We use the BC estimates with eq. 4 and estimate vonB growth model parameters using nls, both weighted by the number of fish at each age (WBC) and unweighted. This nls method includes both the sampling scheme and the first phase information through the weights in eq. 12, but not the individual data. An important disadvantage of this approach is that estimates of σ_a^2 will not fully reflect between-individual variability.

Simulation study

Population simulations were similar to the approach described in Piner et al. (2016) (see Table 2 for details on distributions and models used in simulations). In this paper, all the lengths are in centimetres and ages are in years. Recruitment for each year was modeled using the Beverton–Holt spawner–recruit model, parametrized in terms of steepness, h (e.g., Punt and Cope 2019). Length-at-age was simulated using a vonB model, and weight at age was generated using a simple isometric growth model (Quinn and Deriso 1999). Maturity-at-age was assumed to follow a logistic model (e.g., Jennings et al. 2009), with age at 50% maturity, A_{50}, fixed at log(3)/k, as in Jensen (1996), and age at 95% maturity A_{95} fixed at 4 years from A_{50}. The growth rate k was generated based on the life-history-invariant $k = M / 1.65$ (Charnov 1993). The simulation procedure is illustrated in Fig. 1. Starting from an unfished recruitment $R_0 = \mathrm{e}^9$, we simulated population dynamics for

Fig. 1. Diagram of steps used in population simulations; LSAS is length-stratified age sampling.



61 years with these models and parameter specifications. For each year a new fishing mortality and variation in recruitment were drawn from their corresponding distributions; all other population values were generated once per simulation run. The population in the 61st year was our first phase sample on which LSAS was conducted. The first phase sample size N was not fixed. Length bin sizes were chosen at 1, 2, 3, and 5 cm, as they represented sizes that were frequently observed in practice (e.g., Monnahan et al. 2016). We estimated vonB growth parameters with length at age zero fixed at 3 cm, to compare with the results from Piner et al. (2016), and also estimated vonB parameters with a_0 estimated freely, as in practice this parameter is usually not fixed. For growth parameter estimation using Piner's method, we estimated F as the median F across years within each simulation run, with M fixed at the true population value (see Table 2 for distributions used to derive population parameters), as this method performed well overall in Piner et al. (2016), and in practice F is often estimated from the data while M is fixed at a known value. All other methods did not require an estimate of M.

The purpose of LSAS is to (*i*) control the number of costly age measurements while (*ii*) obtaining length-at-age data evenly spread along the growth curve. Therefore, we assume that the target stratum sample size m_k is proportional to the length bin size so that the total target sample size $\Sigma_k m_k$ does not change with bin size (for (*i*)), and that m_k is independent of k (for (*ii*); i.e., target m_k will be the same, no matter the distribution of length-at-age). To be specific, $m_k = 10 \times$ bin size, for bin sizes = 1, 2, 3, 5. This choice of target stratum sample size is similar to sampling strategies used in real fisheries survey sampling (e.g., American plaice data used in section below on American plaice growth). To demonstrate the effect of a change in the second phase sampling fraction (i.e., N and m_k) on the performance of the estimation approaches, we also conducted a simulation experiment with $m_k = 50$ and $R_0 = e^9$ and with $m_k = 50$ and $R_0 = 5 \times e^9$ for bin size equal to 1 cm and a_0 freely estimated.

This data generation and parameter estimation were repeated 1000 times. The estimation performance was measured using the relative root mean squared error (RRMSE)

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{1000}(\mathrm{est}_i - \mathrm{true}_i)^2}{1000}}, \qquad \mathrm{RRMSE} = \frac{\mathrm{RMSE}}{|\mathrm{true}^*|} \times 100$$

and relative bias (RB)

$$\mathrm{Bias} = \frac{\sum_{i=1}^{1000}(\mathrm{estimate}_i - \mathrm{true}_i)}{1000}, \qquad \mathrm{RB} = \frac{\mathrm{Bias}}{|\mathrm{true}^*|} \times 100$$

Here, true_i denotes the parameter value in the ith simulation, and true^* denotes the mean value of the parameter distribution. The parameters for L_∞, CV, and k were randomly drawn from their corresponding distributions and thus varied across iterations. For the life-history-invariant k, 0.30/1.65 was used for true^*, as 0.30 was the mean of the simulated natural mortality used in the calculation of k.

The estimation was performed in R using Template Model Builder (TMB) (Kristensen et al. 2016), an R package for fast evaluation of the negative log-likelihood function and its gradients, which were then provided to the R function nlminb() for likelihood maximization.

American plaice growth

The two best-fitting methods were used to estimate the vonB growth parameters using a dataset collected by Fisheries and Oceans Canada (DFO) in Northwest Atlantic Fisheries Organization (NAFO) Divisions 3L, 3N, and 3O. We fit the data for the latest year available to us, 2014. The growth models were fit for female plaice only, since male and female plaice follow different growth curves. The actual sampling was conducted at various sites evenly spread across each division to ensure the spatial representativeness of the data, but with an overall sampling goal (m_k) in each division of about 25 age measurements per 2 cm length stratum by sex if length was greater than or equal to 10 cm and about 15 age measurements per stratum without distinguishing sex if length was less than 10 cm. For simplicity we assume that there is no spatial variation in size-at-age of these fish and therefore can neglect the spatial structure of the sampling scheme and regard the whole sample in each division as a LSAS sample (see Appendix C). Spatiotemporal variation in size-at-age is investigated and reported elsewhere.

The sampling in each division was treated as independent, so the likelihood function for the data are given by

$$L_{3\mathrm{LNO}} = L_{3\mathrm{L}} L_{3\mathrm{N}} L_{3\mathrm{O}} \tag{13}$$

where $L_{3\mathrm{L}}$, $L_{3\mathrm{N}}$, and $L_{3\mathrm{O}}$ are the likelihoods of each division. To compare the two best-fitting methods with the methods that are commonly applied in practice, we also estimated vonB parame-

Table 3. RRMSEs (relative root mean squared errors) and RelBias (relative bias) for vonB parameter estimates from 1000 simulations with length at age 0 fixed at 3 cm, for various length bin sizes.

Method	Random	Piner	PF	BC	WBC	HT	CW	Candy	EP
RRMSE									
1 cm									
L_∞	21.3	>100	3.5	11.0	5.2	2.7	2.7	1.7	1.6
k	23.1	>100	5.7	15.1	6.0	3.2	3.2	2.6	2.4
CV	27.0	34.5	3.9	—	—	5.0	5.0	4.1	3.9
2 cm									
L_∞	21.6	>100	3.3	11.2	5.2	3.2	3.2	1.7	1.6
k	23.3	>100	5.7	15.3	6.2	4.0	4.0	2.6	2.5
CV	24.1	35.4	4.0	—	—	5.4	5.4	3.9	3.8
3 cm									
L_∞	21.9	>100	3.3	11.0	5.3	3.6	3.6	1.7	1.6
k	23.6	>100	5.3	15.3	6.1	4.3	4.3	2.6	2.6
CV	21.5	33.9	4.1	—	—	5.7	5.7	3.8	3.8
5 cm									
L_∞	23.2	>100	2.9	11.5	5.5	3.7	3.7	1.8	1.7
k	23.8	>100	4.8	15.9	7.2	4.7	4.7	2.8	2.8
CV	20.6	35.5	4.3	—	—	5.8	5.8	4.0	4.0
RelBias									
1 cm									
L_∞	17.7	−0.5	0.9	7.9	3.3	0.0	0.0	0.1	−0.4
k	−18.9	>100	−1.2	−9.4	−1.5	0.1	0.1	−0.3	0.4
CV	25.7	−7.0	1.1	—	—	2.6	2.6	3.0	2.7
2 cm									
L_∞	18.0	−4.0	0.5	8.1	3.3	0.0	0.0	0.0	−0.3
k	−19.1	>100	−0.5	−9.7	−1.5	0.2	0.2	−0.1	0.3
CV	21.9	−7.6	1.0	—	—	2.5	2.5	2.6	2.4
3 cm									
L_∞	18.3	−7.5	0.5	8.1	3.4	0.2	0.2	−0.0	−0.3
k	−19.3	>100	−0.5	−9.8	−1.6	0.0	0.0	0.0	0.3
CV	18.4	−7.8	1.0	—	—	2.3	2.3	2.4	2.2
5 cm									
L_∞	18.9	>100	0.2	8.6	3.6	0.1	0.2	−0.1	−0.3
k	−19.3	>100	−0.2	−10.3	−1.7	0.0	0.0	0.0	0.2
CV	14.3	−8.4	1.3	—	—	2.4	2.3	2.2	2.1

Note: >100 are results that are greater than 100%.

Table 4. RRMSEs (relative root mean squared errors) and RelBias (relative bias) for vonB parameter estimates from 1000 simulations with a_0 estimated freely, for various length bin sizes.

Method	Random	Piner	PF	BC	WBC	HT	CW	Candy	EP
RRMSE									
1 cm									
L_∞	15.3	16.4	4.2	13.0	7.1	3.5	3.5	2.3	2.2
k	18.0	>100	11.3	21.9	10.9	5.4	5.4	4.4	4.2
CV	28.0	15.8	5.4	—	—	5.6	5.6	6.1	5.8
a_0	>100	>100	>100	>100	>100	43.7	43.7	39.4	37.2
2 cm									
L_∞	14.8	>100	4.0	12.7	7.0	4.0	4.0	2.1	2.1
k	17.0	>100	9.5	22.0	10.9	5.8	5.8	4.2	4.2
CV	23.3	31.9	5.0	—	—	6.2	6.2	5.1	5.0
a_0	>100	>100	>100	>100	>100	56.7	56.7	46.2	44.6
3 cm									
L_∞	14.2	16.3	3.7	12.9	7.2	4.8	4.8	2.1	2.1
k	15.7	>100	8.6	22.3	11.0	6.1	6.2	4.2	4.2
CV	20.1	16.3	4.4	—	—	6.2	6.2	4.6	4.5
a_0	>100	>100	92.0	>100	241.3	62.0	61.8	46.8	47.0
5 cm									
L_∞	13.9	50.9	3.6	13.6	7.3	5.2	5.3	2.1	2.1
k	13.9	>100	8.1	23.0	12.1	7.1	7.1	4.5	4.5
CV	15.8	41.6	4.3	—	—	6.4	6.4	4.2	4.2
a_0	>100	>100	80.3	>100	>100	66.9	66.5	49.0	48.8
RelBias									
1 cm									
L_∞	13.4	−1.3	0.3	8.9	5.1	0.4	0.4	0.0	−0.8
k	−14.6	>100	1.6	−13.7	−6.6	−0.4	−0.4	−0.5	0.9
CV	27.0	1.0	−0.1	—	—	3.6	3.6	4.4	4.0
a_0	10.3	> −100	39.4	> −100	> −100	−5.9	−5.9	−6.9	2.6
2 cm									
L_∞	12.9	>100	0.5	9.0	5.1	0.5	0.5	−0.1	−0.6
k	−13.4	>100	0.2	−13.8	−6.6	−0.4	−0.4	−0.1	0.7
CV	20.8	2.4	0.8	—	—	3.4	3.4	3.4	3.1
a_0	38.9	> −100	11.3	> −100	> −100	−6.0	−6.2	−1.8	3.8
3 cm									
L_∞	12.2	−1.0	0.5	9.1	5.1	0.5	0.4	−0.3	−0.6
k	−12.0	>100	−0.0	−14.0	−6.7	−0.3	−0.3	0.2	0.8
CV	16.4	0.4	1.2	—	—	3.2	3.2	3.0	2.8
a_0	65.1	> −100	3.7	> −100	> −100	−4.4	−4.1	2.4	6.3
5 cm									
L_∞	11.6	1.2	0.3	9.5	5.2	0.7	0.7	−0.2	−0.4
k	−10.0	>100	0.1	−14.7	−6.7	−0.2	−0.2	0.3	0.7
CV	10.0	2.8	1.4	—	—	3.3	3.3	2.5	2.3
a_0	79.4	> −100	1.7	> −100	> −100	−4.5	−4.3	1.9	4.4

Note: >±100 are results that are greater than 100%.

ters using the BC method, both weighted and unweighted, and the random method.

Results

Results for simulation study

The random and Piner methods performed the worst overall. These methods ignored or mis-specified both the sampling scheme and the first phase sample. When a_0 was fixed, Piner's method had the largest RRMSEs for all parameter estimates and for all length bin sizes (see Table 3). The RRMSE for k for Piner's method was much larger than the RRMSE for k for any other method when a_0 was fixed. For example, for the 1 cm length bin, the RRMSE for Piner's method was over 100%; all other methods had RRMSEs for k that were below 25% for that bin size. When a_0 was freely estimated, the RRMSEs for k using Piner's method were again much larger than the RRMSEs for k for all other methods (see Table 4). In terms of relative bias, Piner's method had the largest bias for k, no matter the bin size and whether or not a_0 was estimated freely. Overall, the relative bias of CV estimate was larger for the random method than for any other method.

The bias-corrected mean length-at-age methods, both weighted (WBC) and unweighted (BC), performed slightly better than the random and Piner methods, with RRMSEs that ranged between 5% and 23% for L_∞ and k, for all length bin sizes and whether or not a_0 was estimated freely. The RRMSEs for a_0 were always greater than 100%. The relative bias for the BC and WBC methods were larger than for any of the other methods that included the first phase length information and the sampling scheme (i.e., HT, CW, Candy, and EP).

The PF, HT, and CW methods were the next best-fitting methods. The PF method had the smallest RRMSEs among all the methods that did not include the first phase sampling (i.e., Piner and random), with the RRMSEs decreasing as the length bin size increased. When a_0 was fixed, the HT and CW methods had identical RRMSEs, no matter the bin sizes — these RRMSEs increased slightly as the bin size increased, ranging between 3% and 6%. Overall, the relative bias when a_0 was fixed for both the HT and CW methods was small for all parameter estimates and length bin sizes. When a_0 was freely estimated, the CW and HT methods had almost identical RRMSE values for all parameter estimates and all length bin sizes, with the largest RRMSEs seen for a_0.

Out of all nine methods, EP and Candy's methods performed the best. When a_0 was fixed, the RRMSEs for the EP method were slightly smaller than the RRMSEs for Candy's method for all length bin sizes, with the difference in the RRMSEs decreasing as the bin size increased. Overall, the RRMSEs for both methods when a_0 was fixed were below 4%. When a_0 was freely estimated,

Table 5. RRMSEs (relative root mean squared errors) and RelBias (relative bias) for vonB parameter estimates from 1000 simulations with a_0 estimated freely for 1 cm bin sizes, target stratum sample size m_k = 50, and simulated population size (N) or five times simulated population size ($5N$).

Method	Random	Piner	PF	BC	WBC	HT	CW	Candy	EP
RRMSE									
N; 1 cm; m_k = 50									
L_∞	9.4	11.0	2.2	7.1	2.6	1.9	1.9	1.3	1.3
k	12.5	23.6	5.0	15.3	4.0	2.7	2.7	2.3	2.2
CV	24.3	7.5	3.2	—	—	5.0	5.0	5.4	5.3
a_0	>100	>100	68.0	>100	>100	29.9	29.9	25.5	25.0
$5N$; 1 cm; m_k = 50									
L_∞	14.9	11.1	2.1	9.4	3.3	1.7	1.7	1.2	1.3
k	18.1	25.6	5.4	20.8	5.5	2.6	2.6	2.0	2.0
CV	28.0	7.4	3.0	—	—	4.8	4.8	5.2	5.2
a_0	>100	>100	73.7	>100	> −100	27.8	27.8	22.1	21.9
RelBias									
N; 1 cm; m_k = 50									
L_∞	7.9	0.8	0.2	3.5	1.7	0.2	0.2	−0.1	−0.5
k	−8.7	0.8	0.1	−4.8	−1.2	−0.3	−0.3	−0.2	0.4
CV	23.5	3.8	1.6	—	—	3.8	3.8	4.0	3.9
a_0	14.6	−31.1	11.1	> −100	> −100	−5.5	−5.5	−4.5	−0.6
$5N$; 1 cm; m_k = 50									
L_∞	13.3	0.4	0.2	6.5	2.4	0.0	0.0	−0.5	−0.7
k	−14.9	2.2	−0.0	−12.8	−2.8	−0.1	−0.1	0.3	0.7
CV	27.2	2.8	1.2	—	—	3.5	3.5	4.0	4.0
a_0	18.5	−31.6	14.6	> −100	> −100	−2.5	−2.5	−0.6	1.9

Note: >±100 are results that are greater than 100%.

the RRMSEs were smaller for all parameter estimates for the EP method for the 1 cm length bin, with the differences between the RRMSEs decreasing as the bin size increased. In terms of relative bias, Candy's method had slightly smaller relative bias for L_∞ and k than for the EP method, for all length bin sizes and whether or not a_0 was estimated freely. The relative bias for CV was slightly smaller for the EP method than for Candy's method, for all length bin sizes and whether or not a_0 was estimated freely.

In Table 5, when R_0 remained at e^9, but m_k increased to 50 from 10 in Table 4, the performance of all the approaches including the random method improved with reduced RRMSEs and relative biases. When both R_0 and m_k increased by five times those in Table 4, namely R_0 increase to $5 \times e^9$ and m_k to 50, the RRMSEs and relative biases of the random method did not improve, but the RRMSEs and relative bias of all the other approaches decreased.

American plaice growth

We estimated the vonB parameters for female American plaice in Divisions 3LNO in 2014 using the two best-performing EP and Candy's methods. These methods produced smaller estimates for L_∞ and larger estimates for k than the random and BC methods (see Table 6). The EP method had the smallest standard error for L_∞, at 3.43. The standard errors for k and CV were identical for Candy's, EP, and the random method. The BC methods had the largest estimates for L_∞, at 109.59 for the unweighted method and 89.47 for the weighted method, with the largest standard errors overall for all parameter estimates.

In Fig. 2, it may seem curious that the vonB growth curves estimated using Candy's and the EP methods do not capture the central trend in the data and as such may seem biased. However, it is the LSAS data that are "biased" in a random sample sense. Figure 3 shows the population (N) lengths-at-age for one of our simulated runs and Fig. 4 shows the subsample (n) from that population. The predicted vonB growth curves using the second phase fully observed lengths and ages are also shown on each plot. It is clear from Fig. 4 that the subsample is over-sampling larger, older fish and shorter, younger fish, which is not representative of the population. Figure 3 shows that the EP and Candy's methods are in fact best-fitting, as they capture the real trend in the population and not the LSAS bias in the sample. This LSAS bias in the data must be considered when assessing the validity of model assumption, and Zheng and Cadigan (2019) describe the method to compute residual diagnostics.

Table 6. Parameter estimates and standard errors (SEs) for vonB model parameters for female American plaice data in NAFO Divisions 3LNO in 2014.

Method	L_∞ (SE)	k (SE)	CV (SE)	a_0 (SE)
Random	88.552 (4.942)	0.058 (0.005)	0.123 (0.003)	−0.097 (0.092)
BC	109.589 (13.200)	0.040 (0.008)	—	−0.603 (0.364)
WBC	89.472 (7.620)	0.055 (0.008)	—	−0.025 (0.246)
Candy	80.712 (4.522)	0.062 (0.005)	0.124 (0.003)	−0.223 (0.106)
EP	75.791 (3.427)	0.066 (0.005)	0.117 (0.003)	−0.425 (0.112)

Discussion

Overall, our simulations showed that ignoring the first phase sampling information and mis-specifying or ignoring the sampling scheme had large impacts on the accuracy of von Bertalanffy (vonB) parameter estimates based on length-stratified age samples (LSAS). Both the Piner and random methods had large relative root mean squared error (RRMSE) for all length bin sizes. Incorporating the sampling scheme correctly, as in the partial-full (PF) method, reduced the RRMSE significantly.

The empirical proportion (EP) method performed better than the methods that did not include the first phase information (i.e., random, Piner), and the methods that used the bias-corrected mean length-at-age estimates, both weighted (WBC) and unweighted (BC). The latter point is important because the bias-corrected methods are actively used in fisheries stock assessments to model growth (see, e.g., Brattey et al. 2018). The EP method also performed slightly better than Candy's method and substantially better than the PF method that did not include first phase sampling information. In this simulation study, the first phase samples were very large so that the strata with nonnegligible occupation probability Q_k were mostly nonempty, which reduced the improvement of the EP approach over Candy's method. It would be interesting to investigate the magnitude of the difference in the RRMSEs when the first phase sample sizes are lower, and as a result some of the length strata are empty because we expect the EP method will perform substantially better than Candy's approach in this situation. The full-data likelihood method proposed by Zheng and Cadigan (2019) seems less practically useful than Candy's or the EP method because of its performance and the strong assumptions (i.e., an effective estimation of age distribution) required for the full-data likelihood method.

For almost all methods, when a_0 was freely estimated, the corresponding RRMSEs for a_0 increased as the bin size increased. This was not surprising because in our simulations the abundance of small fish sampled decreased with length bin size, so increasing the length bin size reduced the likelihood that a smaller fish was selected. Thus, for wider bins there were fewer fish sampled that were close to length zero, which provided less information with which to estimate a_0. As a result, when the bin size increased, the estimates of a_0 were less precise, which was reflected in the larger RRMSEs.

Theoretically, we expect Piner's approach to perform better than the random method, since it incorporated some information about the sampling scheme. Piner et al. (2016) also suggested that when the mortality rate Z was properly specified, their method performed better than the random approach for LSAS in terms of relative bias and the distribution of relative bias. Our simulations showed that the relative bias for Piner's method was consistently larger for k and for a_0. Our results also indicate that overall, Piner's method performed the worst in terms of RRMSE for all parameter

Fig. 2. Estimated von Bertalanffy (vonB) growth model for female American plaice data in Divisions 3LNO in 2014. The points have added transparency to better represent the density of the observations. The various fits are as follows: blue line (random), green line (empirical proportion (EP)), yellow line (Candy), purple line (bias-corrected mean length-at age (BC)) and orange line (weighted BC (WBC)). [Colour online.]

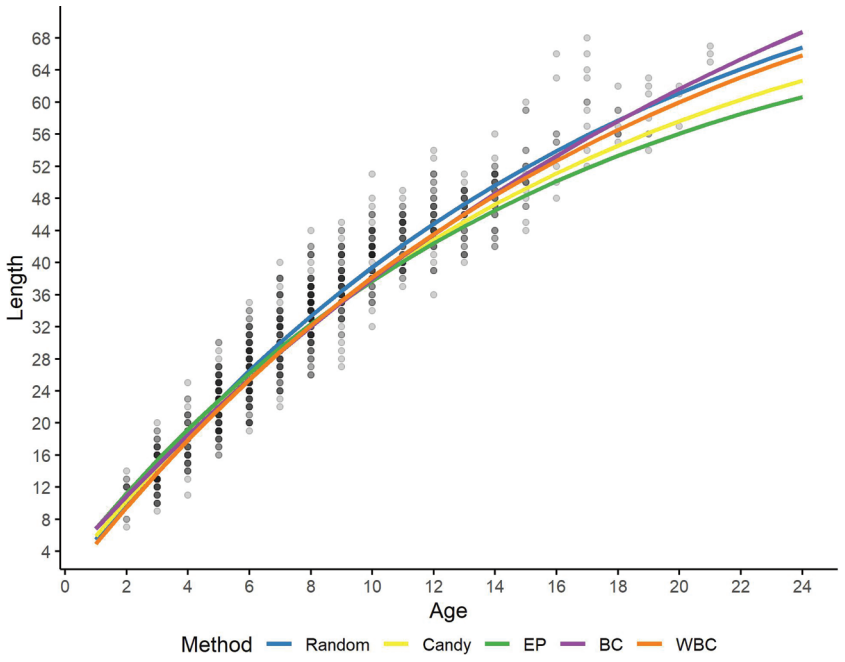
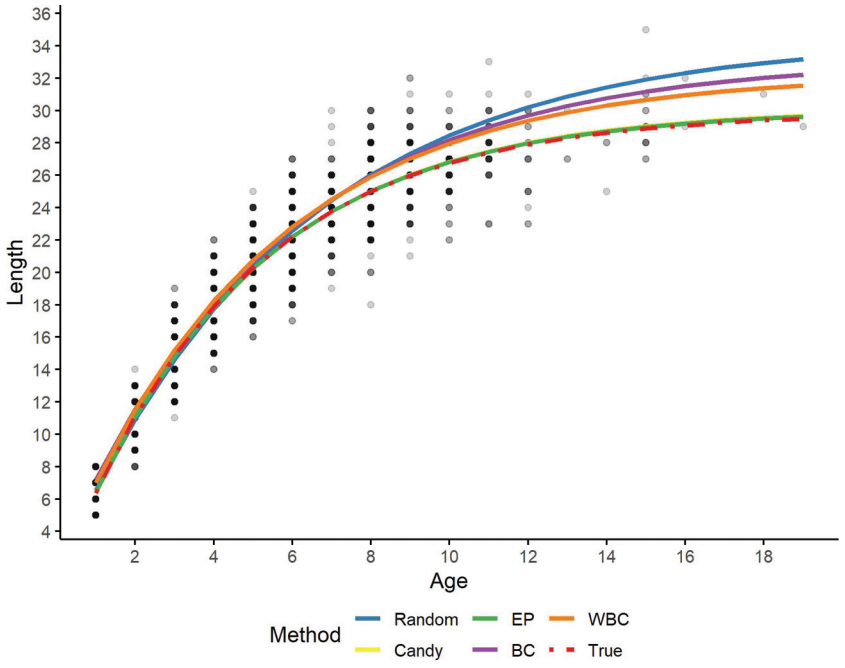


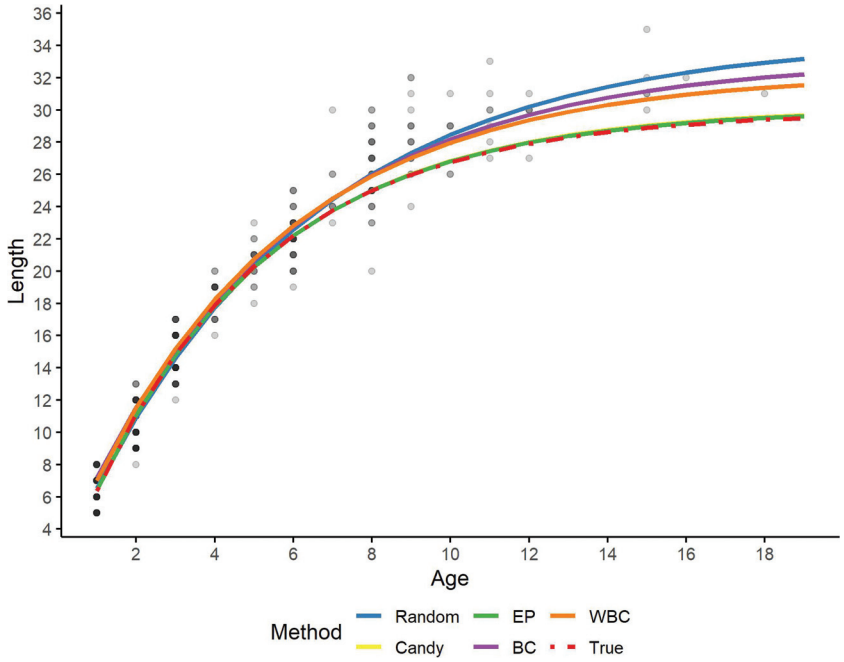
Fig. 3. Plot of simulated population N with estimated vonB growth model from subsample n. The points have added transparency to better represent the density of the observations. The various fits are as follows: blue line (random), green line (EP), yellow line (Candy), purple line (BC), orange line (WBC), and dashed red line (population growth model: True). [Colour online.]



estimates, length bin sizes, and whether or not a_0 was estimated freely, even though an approximately true mortality rate value was used for the Piner estimation. A primary reason for the poor performance of Piner's method was that it adopted an inappropriate model (eq. 8) for the age distribution. In comparison, some of the other approaches, including the random method, used the estimated age distribution or was conditioned on age and hence were not influenced by this issue. Thus, mis-specifying the age distribution had a large impact on the RRMSE. Unless there is strong evidence for a specific age distribution for a population, we would not recommend specifying the age distribution outside the model.

The estimation methods in this paper did not explicitly differentiate between individual variation in growth and length measurement error. All sources of error in the models were treated as one (i.e., confounded), and this is an idealized approach, because in reality there may be various sources of variability. Individuals may have L_∞ and k values that differ from the population mean parameters (Sainsbury 1980; Shelton et al. 2013). As well, age mea-

Fig. 4. Plot of simulated subpopulation n with estimated vonB growth model from subsample. The points have added transparency to better represent the density of the observations. The various fits are as follows: blue line (random), green line (EP), yellow line (Candy), purple line (BC), orange line (WBC), and dashed red line (population growth model: True). [Colour online.]



surement error can occur in practice when conducting age readings on otoliths and can lead to incorrect estimates of growth model parameters (e.g., Cope and Punt 2007; Dey et al. 2019). Not accounting for these sources of error can lead to biased estimates of L_∞ and k. Future research should assess the estimation performance of the methods in this paper when age measurement errors are present.

We also note that although Candy's method performed well, there are two drawbacks to using the probability of length conditional on age. First, it is impossible to address age measurement errors with a structural-errors in variables (SEV) approach (e.g., Carroll et al. 2006; Cope and Punt 2007; Dey et al. 2019). As noted above, the effect of ignoring age measurement errors is well known and can lead to incorrect estimates of growth parameters and in particular underestimation of L_∞ and overestimation of k because of the well-known bias-attenuation problem with covariate measurement errors. Such biases will tend to lead to overestimation of F_{msy} and underestimation of B_{msy} and consequently suboptimal and possibly unsustainable harvest advice. Second, Candy's method cannot use the first phase length data with the likelihood function as constructed in Zheng and Cadigan (2019). Because of these drawbacks, we suggest that the EP method may be easier to implement and more flexible in practice. Also because of these considerations, we tried to incorporate $P(a)$ in all the approaches except Candy, even though under the current simulation setup, conditional-on-age approach gives the same growth parameter estimates for random, HT, and CW approaches, since $P(a)$ involves no growth model parameters.

Increasing sample size leads to improved estimates if an estimator is unbiased. However, the performance of a biased estimator may not improve with sample size. With LSAS, if the first phase stratum sample sizes $N_{k(l)}$ and target second phase stratum sample size m_k are all increased by five times for example, the bias in the data introduced by LSAS is not alleviated, and the performance of the estimation approaches neglecting LSAS do not improve. This is why when both R_0 and m_k increased by five times in Table 5 compared with Table 4, the performance of the random method did not improve as the other approaches did. Note that when the bin size equals 1 cm, Piner's method is in fact the conditional approach discussed in Hausman and Wise (1982), Breslow and Cain (1988), Pfeffermann and Sverchkov (1999), and Scott and Wild (2011), which gives some account for the sampling strategy; hence, its performance also improved in this case. On the other hand, when the first phase sample size remained unchanged (equivalently R_0 remained at e^9), but m_k increased from 10 to 50, many strata became fully sampled or close to fully sampled for ageing, the bias in LSAS data was alleviated, and the performance of random method also improved in Table 5. In the extreme case, when all first phase fish are selected for ageing, then the data become a simple random sample from the population, and the random method will work well in this case.

The bias from mis-specifying the three components (i.e., first phase sample, second phase sample, and sampling design) is not limited to data collected for growth parameter estimation. The stratified sampling design is often used to select subsamples of fish to estimate important biological parameters such as length-at-maturity or length–fecundity relationships (Quinn and Deriso 1999). We suggest that the three components should not be neglected for any samples collected using stratified sampling and that parameter estimates will be biased if the sampling design is ignored. Further work should be done to investigate the magnitude of these biases for similarly collected data.

In conclusion, we demonstrate the importance of incorporating the first phase length information and the sampling design when estimating growth parameters with length-stratified age samples. Correctly accounting for the sampling scheme can greatly reduce the RRMSE for all parameter estimates. Incorporating the information from the first phase sample in combination with the sampling scheme reduced the RRMSE. We also showed that improperly modeling the age distribution can lead to large misspecifications of growth parameter estimates, and using the LSAS bias-corrected mean length-at-age can also produce less reliable parameter estimates. These growth parameter estimates are often incorporated into stock assessments to assess the overall health of a fish population. If the parameter estimates are incorrect, then

we may be under- or overestimating the health of a fish stock, which can lead to incorrect management and stock advice.

Acknowledgements

Research funding to AP, NZ, and NC was provided by the Ocean Frontier Institute, through an award from the Canada First Research Excellence Fund. Research funding to NC was also provided by the Ocean Choice International Industry Research Chair program at the Marine Institute of Memorial University of Newfoundland. Many thanks for the comments from the associate editor and two anonymous reviewers that greatly improved this manuscript.

References

Agresti, A. 2003. Categorical data analysis. Vol. 482. John Wiley & Sons.

Berg, C.W., and Kristensen, K. 2012. Spatial age–length key modelling using continuation ratio logits. Fish. Res. **129**: 119–126. doi:10.1016/j.fishres.2012.06.016.

Bettoli, P.W., and Miranda, L.E. 2001. Cautionary note about estimating mean length at age with subsampled data. North Am. J. Fish. Manage. **21**(2): 425–428. doi:10.1577/1548-8675(2001)021<0425:CNAEML>2.0.CO;2.

Brattey, J., Cadigan, N., Dwyer, K., Healy, B., Ings, D., Lee, E., et al. 2018. Assessment of the northern cod (*Gadus morhua*) stock in NAFO divisions 2J3KL in 2016. DFO Can. Sci. Advis. Sec. Res. Doc. 2018/018.

Breslow, N.E., and Cain, K.C. 1988. Logistic regression for two-stage case-control data. Biometrika, **75**(1): 11–20. doi:10.1093/biomet/75.1.11.

Breslow, N.E., Lumley, T., Ballantyne, C.M., Chambless, L.E., and Kulich, M. 2009. Improved Horvitz–Thompson estimation of model parameters from two-phase stratified samples: applications in epidemiology. Stat. Biosci. **1**(1): 32–49. doi:10.1007/s12561-009-9001-6.

Cadigan, N.G. 2016. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. Can. J. Fish. Aquat. Sci. **73**(2): 296–308. doi:10.1139/cjfas-2015-0047.

Candy, S.G., Constable, A.J., Lamb, T., and Williams, R. 2007. A von Bertalanffy growth model for toothfish at Heard Island fitted to length-at-age data and compared to observed growth from mark–recapture studies. CCAMLR Science, **14**: 43–66.

Carroll, R.J., Ruppert, D., Stefanski, L.A., and Crainiceanu, C.M. 2006. Measurement error in nonlinear models: a modern perspective. CRC press.

Charnov, E.L. 1993. Life history invariants: some explorations of symmetry in evolutionary ecology. Vol. 6. Oxford University Press, USA.

Chen, Y., Jackson, D., and Harvey, H. 1992. A comparison of von Bertalanffy and polynomial functions in modelling fish growth data. Can. J. Fish. Aquat. Sci. **49**(6): 1228–1235. doi:10.1139/f92-138.

Cope, J.M., and Punt, A.E. 2007. Admitting ageing error when fitting growth curves: an example using the von Bertalanffy growth function with random effects. Can. J. Fish. Aquat. Sci. **64**(2): 205–218. doi:10.1139/f06-179.

Deville, J.-C., and Särndal, C.-E. 1992. Calibration estimators in survey sampling. J. Am. Stat. Assoc. **87**(418): 376–382. doi:10.1080/01621459.1992.10475217.

Dey, R., Cadigan, N., and Zheng, N. 2019. Estimation of the von Bertalanffy growth model when ages are measured with error. J. R. Stat. Soc.: Ser. C, Appl. Stat. **68**: 1131–1147. doi:10.1111/rssc.12340.

Doubleday, W. 1981. Bottom trawl surveys. Can. Spec. Pub. Fish. Aquat. Sci. **58**: 1–273.

Echave, K.B., Hanselman, D.H., Adkison, M.D., and Sigler, M.F. 2012. Interdecadal change in growth of sablefish (*Anoplopoma fimbria*) in the Northeast Pacific ocean. Fish. Bull. **110**(3): 361–374.

Francis, R.C. 2016. Growth in age-structured stock assessment models. Fish. Res. **180**: 77–86. doi:10.1016/j.fishres.2015.02.018.

Hausman, J.A., and Wise, D.A. 1982. Stratification on endogenous variables and estimation: The gray income maintenance experiment. *In* Structural Analysis of Discrete Data: With Econometric Applications. *Edited by* C. Manski and D. McFadden. M.I.T. Press. Cambridge, Mass. Chapter 10, pp. 365–391.

Hilborn, R., and Walters, C.J. 2013. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Springer Science & Business Media.

Hoggarth, D.D. 2006. Stock assessment for fishery management: A framework guide to the stock assessment tools of the fisheries management and science programme. Vol. 487. Food & Agriculture Org.

Hsieh, D.A., Manski, C.F., and McFadden, D. 1985. Estimation of response probabilities from augmented retrospective observations. J. Am. Stat. Assoc. **80**(391): 651–662. doi:10.1080/01621459.1985.10478165.

Jennings, S., Kaiser, M., and Reynolds, J.D. 2009. Marine fisheries ecology. John Wiley & Sons.

Jensen, A. 1996. Beverton and Holt life history invariants result from optimal trade-off of reproduction and survival. Can. J. Fish. Aquat. Sci. **53**(4): 820–822. doi:10.1139/f95-233.

Kalbfleisch, J.D., and Lawless, J.F. 1988. Likelihood analysis of multi-state models for disease incidence and mortality. Stat. Med. **7**(1-2): 149–160. doi:10.1002/sim.4780070116.

Kimura, D.K. 2008. Extending the von Bertalanffy growth model using explanatory variables. Can. J. Fish. Aquat. Sci. **65**(9): 1879–1891. doi:10.1139/F08-091.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B.M. 2016. TMB: Automatic differentiation and Laplace approximation. J. Stat. Softw. **70**(5). doi:10.18637/jss.v070.i05.

Lee, H., Thomas, L., Piner, K., and Maunder, M. 2017. Effects of age-based movement on the estimation of growth assuming random-at-age or random-at-length data. J. Fish Biol. **90**(1): 222–235. doi:10.1111/jfb.13177.

Maunder, M.N., and Piner, K.R. 2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. **72**(1): 7–18. doi:10.1093/icesjms/fsu015.

Methot, R.D., and Wetzel, C.R. 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. **142**: 86–99. doi:10.1016/j.fishres.2012.10.012.

Monnahan, C.C., Ono, K., Anderson, S.C., Rudd, M.B., Hicks, A.C., Hurtado-Ferro, F., et al. 2016. The effect of length bin width on growth estimation in integrated age-structured stock assessments. Fish. Res. **180**: 103–112. doi:10.1016/j.fishres.2015.11.002.

Neyman, J. 1938. Contribution to the theory of sampling human populations. J. Am. Stat. Assoc. **33**(201): 101–116. doi:10.1080/01621459.1938.10503378.

Pfeffermann, D., and Sverchkov, M. 1999. Parametric and semi-parametric estimation of regression models fitted to survey data. Sankhyā: Indian J. Stat. Ser. B: 166–186.

Piner, K.R., Lee, H.-H., and Maunder, M.N. 2016. Evaluation of using random-at-length observations and an equilibrium approximation of the population age structure in fitting the von Bertalanffy growth function. Fish. Res. **180**: 128–137. doi:10.1016/j.fishres.2015.05.024.

Punt, A.E., and Cope, J.M. 2019. Extending integrated stock assessment models to use non-depensatory three-parameter stock–recruitment relationships. Fish. Res. **217**: 46–57. doi:10.1016/j.fishres.2017.07.007.

Quinn, T.J., and Deriso, R.B. 1999. Quantitative fish dynamics. Oxford University Press.

Quist, M.C., Pegg, M.A., and DeVries, D.R. 2012. Age and growth. Fisheries techniques, 3rd edition. American Fisheries Society, Bethesda, Maryland. pp. 677–731.

Saegusa, T., and Wellner, J.A. 2013. Weighted likelihood estimation under two-phase sampling. Ann. Stat. **41**(1): 269–295. doi:10.1214/12-AOS1073.

Sainsbury, K. 1980. Effect of individual variability on the von Bertalanffy growth equation. Can. J. Fish. Aquat. Sci. **37**(2): 241–247. doi:10.1139/f80-031.

Schueller, A.M., Williams, E.H., and Cheshire, R.T. 2014. A proposed, tested, and applied adjustment to account for bias in growth parameter estimates due to selectivity. Fish. Res. **158**: 26–39. doi:10.1016/j.fishres.2013.10.023.

Scott, A.J., and Wild, C.J. 2011. Fitting regression models with response-biased samples. Can. J. Stat. **39**(3): 519–536.

Shackell, N.L., Frank, K.T., Fisher, J.A., Petrie, B., and Leggett, W.C. 2010. Decline in top predator body size and changing climate alter trophic structure in an oceanic ecosystem. Proc. R. Soc. B: Biol Sci. **277**(1686): 1353–1360. doi:10.1098/rspb.2009.1020.

Shelton, A.O., Satterthwaite, W.H., Beakes, M.P., Munch, S.B., Sogard, S.M., and Mangel, M. 2013. Separating intrinsic and environmental contributions to growth and their population consequences. Am. Nat. **181**(6): 799–814. doi:10.1086/670198.

Smith, S.J., Hunt, J.J., and Rivard, D. (eds.). 1993. Risk Evaluation and Biological Reference Points for Fisheries Management. Special Publication of Fisheries and Aquatic Sciences No. 120. NRC Research Press, Ottawa, Ont., Canada.

Taylor, N.G., Walters, C.J., and Martell, S.J. 2005. A new likelihood for simultaneously estimating von Bertalanffy growth parameters, gear selectivity, and natural and fishing mortality. Can. J. Fish. Aquat. Sci. **62**(1): 215–223. doi:10.1139/f04-189.

Thompson, S. 2012. Sampling. John Wiley & Sons, Hoboken, N.J., USA.

Troynikov, V.S. 1999. Use of Bayes theorem to correct size-specific sampling bias in growth data. Bull. Math. Biol. **61**(2): 355–363. doi:10.1006/bulm.1998.0088.

Vincenzi, S., Crivelli, A.J., Munch, S., Skaug, H.J., and Mangel, M. 2016. Trade-offs between accuracy and interpretability in von Bertalanffy random-effects models of growth. Ecol. Appl. **26**(5): 1535–1552. doi:10.1890/15-1177.

von Bertalanffy, L. 1938. A quantitative theory of organic growth (inquiries on growth laws. II). Human Biol. **10**(2): 181–213.

Zheng, N., and Cadigan, N. 2019. Likelihood methods for basic stratified sampling, with application to von bertalanffy growth model estimation. Open J. Stat. **9**: 623–642. doi:10.4236/ojs.2019.96040.

Appendix A

Without loss of generality, we assume that $l \in S_k$. It follows from conditional probability that

$$f(l, a \mid \text{LSAS}, \boldsymbol{\theta}) = \sum\nolimits_k [f(l, a \mid l \in S_k; \text{LSAS}, \boldsymbol{\theta}) Q_k(\text{LSAS}, \boldsymbol{\theta})] = f(l, a \mid l \in S_k; \text{LSAS}, \boldsymbol{\theta}) Q_k(\text{LSAS}, \boldsymbol{\theta})$$

where $Q_k(\text{LSAS}, \boldsymbol{\theta})$ is the probability for a second phase individual to be in length stratum S_k. Since the subsampling in each stratum is random, $f(l, a \mid l \in S_k; \text{LSAS}, \boldsymbol{\theta}) = f(l, a \mid l \in S_k; \boldsymbol{\theta})$, we can write

$$f(l, a \mid \text{LSAS}, \boldsymbol{\theta}) = f(l, a \mid l \in S_k; \boldsymbol{\theta}) Q_k(\text{LSAS}, \boldsymbol{\theta}) = \frac{f(l, a \mid \boldsymbol{\theta})}{Q_k(\boldsymbol{\theta})} Q_k(\text{LSAS}, \boldsymbol{\theta})$$

Appendix B

For the calibrated weighted (CW) method, weights are defined as $w_i = g_i[N_{k(l)} / n_{k(l)}]$ for $i = 1, \ldots, N$, where $k(l)$ represents length l in stratum k, such that the difference between w_i and $N_{k(l)} / n_{k(l)}$ is minimized subject to some constraints. To calibrate the first-order moment (the mean), we calculate the weights w_i by minimizing the χ^2 distance $G(w_i, d_i) = (w_i - d_i)^2 / 2d_i$, subject to

$$\sum_{i=1}^{N} \gamma_i w_i l_i = \sum_{i=1}^{N} l_i$$

Here, γ_i is a binary indicator of whether or not the ith subject is fully observed. These new calibrated weights can be found using the method of Lagrange multipliers and can easily be implemented using eq. 9 from Breslow et al. (2009). From that equation it follows that

$$g_i = 1 - \lambda_c l_i$$

where

$$\lambda_c = \frac{\sum_{i=1}^{n} \frac{N_{k(l)}}{n_{k(l)}} l_i - \sum_{i=1}^{N} l_i}{\sum_{i=1}^{n} l_i^2}$$

then

$$w_i = (1 - \lambda_c l_i) \frac{N_{k(l)}}{n_{k(l)}}$$

Appendix C

Assume that from a model M, we generated a random sample A with the individuals enumerated as $a_1, a_2, \ldots, a_N$. In an arbitrary length stratum k, there are N_k individuals of sample A that are denoted as $a_{k,1}, a_{k,2}, \ldots, a_{k,N_k}$. Now we make a random permutation of sample A, which gives sample B with individuals $b_1, b_2, \ldots, b_N$, and the N_k individuals in stratum k are now denoted as $b_{k,1}, b_{k,2}, \ldots, b_{k,N_k}$. The individuals in sample B are allocated into N_s survey stations, and length-stratified random sampling is conducted in each stratum with different rules. The final result is to obtain n_k individuals in stratum k for age measurement. Owing to the random permutation between A and B, despite the complicated sampling design, this is in fact a random sample of size n_k from $a_{k,1}, a_{k,2}, \ldots, a_{k,N_k}$ for any stratum k, then all the inference approaches and conclusions of this paper can be applied. Therefore, we can treat the final length-at-age data from sample B (or equivalently from sample A, since A and B have identical set of individuals) as from standard LSAS defined by eq. 2. Equivalent to A, sample B is also a random sample generated from model M. We reach the conclusion that for any random sample generated from model M, we can neglect the complicated station-wise sampling design and treat the final data as from standard LSAS.

Reference

Breslow, N.E., Lumley, T., Ballantyne, C.M., Chambless, L.E., and Kulich, M. 2009. Improved Horvitz–Thompson estimation of model parameters from two-phase stratified samples: applications in epidemiology. Stat. Biosci. **1**(1): 32–49. doi:10.1007/s12561-009-9001-6. PMID:20174455.

4.2 Paper 2

A state-space stock assessment model for American plaice on the Grand Bank of Newfoundland

Andrea M.J. Perreault

Centre for Fisheries Ecosystems Research, Fisheries and Marine Institute of Memorial University,

St. John's, NL, Canada, A1C 5R3

andrea.perreault@mi.mun.ca

Laura Wheeland

Northwest Atlantic Fisheries Centre, Fisheries and Oceans Canada,

PO Box 5667, St John's, NL, Canada A1C 5X1

laura.wheeland@dfo-mpo.gc.ca

M. Joanne Morgan

Northwest Atlantic Fisheries Centre, Fisheries and Oceans Canada,

PO Box 5667, St John's, NL, Canada A1C 5X1

joanne.morgan@dfo-mpo.gc.ca

Noel G. Cadigan

Centre for Fisheries Ecosystems Research, Fisheries and Marine Institute of Memorial University,

St. John's, NL, Canada, A1C 5R3

noel.cadigan@mi.mun.ca

Abstract

The current stock assessment model for American plaice (*Hippoglossoides platessoides*) on the Grand Bank of Newfoundland is a virtual population analysis (VPA). This model does not account for the considerable uncertainty about the landings data for this stock. Retrospective patterns have also been noted in the current assessment with overestimation of spawning stock biomass (SSB) and underestimation of fishing mortality (F). In this work, we develop an aged-based state-space stock assessment model (SSM) for Grand Bank (NAFO Divisions 3LNO) American plaice that accounts for the uncertainties in the landings data and reduces the problem of retrospective patterns. Our SSM fit the data well, with overall trends in SSB and average F (ages 9-14) similar to those estimated from the current VPA. The retrospective patterns for the SSM were reduced for SSB and greatly reduced for average F which should lead to the provision of better scientific advice for the management of this

stock. This model is also a valuable first step in improving our understanding of the stock of American plaice as the flexibility of state-space models are an ideal foundation from which to build more complex and realistic models.

Introduction

American plaice (*Hippoglossoides platessoides*) on the Grand Bank of Newfoundland (NAFO Divisions 3LNO) supported an important commercial fishery historically, accounting for over ten percent of the Canadian groundfish fishery in the 1950's (DFO, 2011). The population size declined rapidly in the 1980's due mostly to overfishing and, although there has been no directed commercial fishing since 1994, there has since been little improvement in the state of the population (*see e.g.* Wheeland, 2018). The major factor that has been attributed to the lack of recovery is overfishing, which has occurred mainly through bycatch in the yellowtail flounder, skate, redfish, and Greenland halibut fisheries (Shelton and Morgan, 2005). It has also been suggested that an increase in the natural mortality rate due to changing ocean temperatures may also be contributing to the lack of recovery (Rideout *et al.*, 2008).

The current stock assessment model for American plaice relies on the use of commercial catch-at-age data. When these data are available, the most commonly applied stock assessment models are virtual population analysis, statistical catch-at-age, and, more recently, integrated and/or state-space models. Virtual population analysis assumes that the catch-at-age data are known without error (Megrey, 1989), whereas statistical catch-at-age models allow for errors in the catch-at-age data, although these models commonly assume that the fisheries' selectivity pattern is constant over blocks of years (Hilbon and Walters, 1992). Integrated models were first introduced in the early 1980's (Fournier and Archibald, 1982) and aim to use as much data as possible in as raw a form as possible, while state-space models include both random errors in the underlying population dynamics model (i.e. for population abundance and fishing mortality rates) and measurement errors in the data (*see e.g.* Nielsen and Berg, 2014; Cadigan, 2015; Albertsen *et al.*, 2016). Advances in computing power and technology have led to an increase in the application of both integrated models (Maunder and Punt, 2013) and state-space models (Aeberhard *et al.*, 2018) as the availability and resolution of data has increased and it is now possible to efficiently integrate out random effects from complex joint likelihood functions.

The current stock assessment model for Grand Bank American plaice is a virtual population analysis that was introduced in the late 90's. This model is informed by catch-at-age data that are derived in part from landings estimates and does not account for the considerable uncertainty about the landings data (Wheeland *et al*., 2018).

Sources of uncertainty include landings estimated from "unspecified flounder" by some countries in the earliest years of available data (*see e.g.* Pitt, 1972) and an increase in foreign catch outside the 200 mile economic exclusive zone in the mid-80's (*e.g.* Brodie, 1986). More recently, the loss of availability of scientific observer data in the NAFO Regulatory area has resulted in various methods applied to obtain landings estimates, including effort ratios and daily catch records (Dwyer et *al*., 2016). As a result, there may be large errors in the landings data and a stock assessment model that incorporates uncertainty in these data may therefore provide a better assessment of the stock.

Another issue that has been noted in the current assessment for American plaice are retrospective patterns, which are consistent directional changes in estimates of stock size as years of data are removed from the assessment model (Mohn, 1999). Retrospective patterns are caused by changes in the accuracy of the data over time and/or spatial and time-varying population processes that are unaccounted for or mis-specified in the model (*see e.g.* Legault, 2009). Systematic retrospective patterns can lead to poor management advice as important population processes (*e.g.* biomass and fishing mortality) may be over- or under-estimated and can result in unsustainable or sub-optimal harvesting advice (Szuwalski *et al*., 2017). To promote sustainable management advice for American plaice on the Grand Bank of Newfoundland, a stock assessment model that reduces or eliminates retrospective patterns is valuable. In this paper, we develop an aged-based state-space stock assessment model for Grand Bank American plaice that accounts for uncertainties in the landings data and reduces the problem of retrospective patterns.

Materials and Methods

There are two components to a state-space stock assessment model: the process model and the observation model. For our application to Grand Bank American plaice, the process model describes how the state of the unobserved fish stock abundance and fishing mortality rates at a given time depend on previous states. The observation model describes how the survey and commercial data depend on the unobserved states (*see e.g.* Aeberhard *et al.*, 2018).

Process Model

The model runs for the years $y = 1960, \ldots, 2017$ for ages $a = 1, \ldots, 15^{+}$, where 15^{+} represents the oldest ages grouped together from ages 15 onwards, called the plus group. For simplicity, we will refer to model ages $a = 1, \ldots, A^{+}$, and years $y = 1, \ldots, Y$. The process model describes how the abundance at age a in year y (i.e. $N_{y,a}$) and

the fishing mortality, $F_{y,a}$ change over time. The $N_{y,a}$ for all ages and years are treated as random effects, with the cohort abundance model modelled as

$$\begin{aligned} \log(N_{y,a}) &= \log(N_{y-1,a-1}) - Z_{y-1,a-1} + \gamma_{y,a} \\ \log(N_{y,A^+}) &= \log\left[N_{y-1,A^+-1}\exp^{-Z_{y-1,A^+-1}} + N_{y-1,A^+}\exp^{-Z_{y-1,A^+}}\right] + \gamma_{y,a}, \end{aligned} \quad (1.1)$$

where $Z_{y,a} = M_{y,a} + F_{y,a}$ is the total mortality rate given by the sum of the natural mortality rate, $M_{y,a}$ (i.e. all mortality unrelated to fishing) and $F_{y,a}$. Here, $M_{y,a}$ is assumed to be known and fixed at 0.50 for ages 1-3, 0.30 for age 4 and 0.20 for all ages 5 and above, except during 1989 to 1996, where it is fixed at 0.53 for all ages 5 and above, as recommended by Morgan and Brodie (2001), 0.83 for ages 1-3 and 0.63 for age 4. We note that this formulation for $M_{y,a}$ is identical to the formulation for the most recent stock assessment model for Grand Bank American plaice. $F_{y,a}$ is set to zero for ages 1-4, as there is no reported catch at these ages. The $\gamma_{y,a}$ are the process errors, assumed to be independent and normally distributed with variance σ_{pe}^2 to be estimated. The numbers at the first ages $N_{y,1}$ are modelled as

$$\log(N_{y,1}) = \mu_{R_y} + \delta_{R_y}, \quad (1.2)$$

where $\mu_{R_y} = \mu_{R_1}$ for $y \leq 1993$ and $\mu_{R_y} = \mu_{R_2}$ for $y > 1993$, and the two mean recruitment parameters $\mu_{R_1}, \mu_{R_2} \in (-\infty, \infty)$ account for the large differences in recruitment between the two time periods and are fixed effect parameters to be estimated. The deviations from the mean recruitment δ_{R_y} are assumed to follow a normal distribution with AR(1) correlation across years, with the AR parameters σ_R^2 and ϕ_R to be estimated, as we expect recruitment to be more alike in years that are closer together.

The fishing mortality rates are modelled similarly,

$$\log(F_{y,a}) = \mu_{F_{y,a}} + \delta_{F_{y,a}}, \quad (1.3)$$

where $\mu_{F_{y,a}}$ is the mean fishing mortality rate and $\delta_{F_{y,a}}$ is the deviation from the mean at each age and year. A separate $\mu_{F_{y,a}}$ is estimated for ages 5, 6, 7, 8, 9 and 10+ (where 10+ represents ages 10-15+) for two blocks: 1960-1994 and 1995-2017 (i.e. twelve fixed effect *F* parameters). The age blocking of the $\mu_{F_{y,a}}$'s were chosen via model building to reflect overall fishery selectivity patterns, and the year blocks were chosen to account for the closure of the commercial fishery in 1994. The $\delta_{F_{y,a}}{}'s$ are treated as random effects and are assumed to follow a normal

distribution, with the deviations at the first age, $\delta_{F_{y,5}}$ assumed to have AR(1) correlation across years, independent from ages 6-15^{+}, with parameters $\sigma^2_{F_5}$, $\phi_{F_5}$to be estimated. We treat the $\delta_{F_{y,a}}{}'s$ separately for age 5 fish as preliminary analyses indicated that trends in *F*'s differed at age 5 compared to older ages. The *F*-deviations at ages 6-15^{+} were treated as a correlated AR(1) process across ages and years, with parameters $\sigma^2_{F6^+}$, $\phi_{F_{A6^+}}$, $\phi_{F_{Y6^+}}$to be estimated. We fit an AR(1) process across ages and years for age 6-15^{+} fish as fish that are closer in age and time are expected to have *F*-deviations that are more similar than those that are further apart.

Observation model

The observation model includes data from the commercial fishery and scientific research trawl surveys. There are two basic types of fishery information: total landed weight, and the size (length, weight) and age composition of the landings. Both these sources of information are used to derive annual fishery catch numbers-at-age. In the integrated assessment model philosophy, these data sources should enter into the assessment model fitting via separate observation models. We particularly want to focus our model estimation to include uncertainty in landings. Therefore, for pragmatic reasons, we used landings information (1960-2017) and the catch proportions-at-age (ages 5-15+ during 1960-2017) as independent data sources for model estimation. Stock size age-based indices are derived from the Canadian fall and spring research surveys in NAFO Divs. 3LNO (*see* Dwyer *et al.*, 2014 for details) and the Spanish research survey in the portions of NAFO Divs. 3NO outside of the Canadian Exclusive Economic Zone (EEZ) (González-Troncoso1 *et al.*, 2017) were also used in model estimation. Indices were for ages 1-15+ for all surveys, for years 1990-2017 for the fall survey (2004 and 2014 omitted due to poor survey coverage), 1985-2016 for the spring survey (2006 and 2015 omitted due to poor survey coverage) and 1997-2016 for the Spanish survey. The Baranov catch equation is used to model commercial catch as a function of $N, F$and Z,

$$C_{y,a} = \frac{F_{y,a}}{Z_{y,a}}(1 - \exp^{-Z_{y,a}})N_{y,a}. \quad (1.4)$$

Model predicted catch proportion at age ($P_a = C_a / \sum_a C_a$) were fit to observed proportions, as described in the next section. Commercial average weights-at-age ($W_{y,a}$) were calculated by Rivard's method (Rivard, 1980) and are used to calculate model predicted landings each year, $L_y = \sum_a W_{y,a} C_{y,a}$.

Age composition data

We fit the age composition data using the continuation-ratio logit (crl) transformation (*see e.g.* Cadigan, 2015; Berg and Kristensen, 2012; Agresti, 2003). A direct observation model for the matrix of observed catch

proportions each year is complicated because $P_{oa} \geq 0$ and $\sum P_{oa} = 1$. We use the crl which maps P_a for $a = 1, \ldots, A_{max}$ into $X_a \in (-\infty, \infty)$ for $a = 1, \ldots, A_{max} - 1$. The unconstrained crls are derived from the multiplicative logistic transformation,

$$X_a = \log\left[\frac{P_a}{P_{a+1}+\ldots+P_{A_{max}}}\right], a = 5, \ldots, A_{max} - 1. \quad (1.5)$$

where A_{max} is the plus group. The inverse transformation of (1.5) is

$$P_a = \begin{cases} \frac{\exp(X_a)}{\prod_{i=1}^{a}(1+\exp(X_i))}, & a = 5, \ldots, A_{max} - 1 \\ \frac{1}{\prod_{i=1}^{A_{max}-1}(1+\exp(X_i))}, & a = A_{max}. \end{cases} \quad (1.6)$$

The crls for the observed catch proportions-at-age data (i.e. $X_{oy,a}$) are calculated from (1.5) and the observation model we use is based on assuming the model residuals ($X_{oy,a} - X_{y,a}$) have a normal distribution with AR(1) correlation across ages and years with parameters ϕ_{C_A}, ϕ_{C_Y} to be estimated, as we expect the crl errors to be similar for fish that are closer in age and time. We fit two variance parameters (σ^2_{C5} , σ^2_{C6+}; one for age 5 fish, and one for ages 6+ fish) as preliminary data analysis indicated that the error variance differed at age 5, the youngest commercially caught age.

Landings data

Dwyer *et al. (*2016) reported uncertainties about the reliability of the landings data for Grand Bank American plaice. To account for this, we treat reported landings as a lower bound for true landings (i.e. not all catches are reported). We assume that there is an upper bound for landings that varies with the reliability of data (see Table 1 for details). We assume the true landings could be accurately estimated with a CV of 2%. Let B_{ly} and B_{uy} denote the lower and upper bounds and $\sigma_C = 0.02$. The observation equation (*i.e.* negative loglikelihood, nll) for the landings bounds data is

$$\text{nll}(L_1, \ldots, L_Y | \{B_{ly}, B_{uy}; y = 1, \ldots Y\}|) = \sum_{y=1}^{Y} \log\left[\Phi\left\{\frac{\log(B_{uy}/L_y)}{\sigma_C}\right\} - \Phi\left\{\frac{\log(B_{ly}/L_y)}{\sigma_C}\right\}\right], \quad (1.7)$$

where $L_1, \ldots, L_Y$ are the model predicted landings. We fixed σ_L at a small value to ensure that the estimates of landings are between the bounds for most years.

Survey data

The model-predicted catch for survey *s* is

$$I_{s,y,a} = q_{s,a} N_{y,a} \exp^{-f_{s,y} Z_{y,a}} \quad (1.8)$$

where f represents the fraction of the year the survey takes place (0.460 for the Canadian spring and Spanish surveys and 0.875 for the Canadian fall survey), and $q_{s,a}$ are the survey catchabilities that are free parameters to be estimated. The q parameters are fit by survey with a separate q estimated for ages 1-9 and grouped for ages 10^{+}, as preliminary model fitting resulted in similar q parameters at the oldest ages and this reduced the number of parameters to be estimated. For the spring and fall surveys, fish of ages 1-4 are given a separate q for each gear period due to issues in conversion from the Engel to the Campelen survey trawl (Dywer *et. al.*, 2016). Diagnostic model fitting found little difference in the q estimates for the Spanish survey for the two gear types at the youngest ages, therefore only one q was estimated at each age. The indices are assumed to follow a normal distribution, with mean $I_{s,y,a}$ and standard deviation $\sigma_{s,a} = \text{cv}_s \cdot I_{s,y,a}$, where cv_s represents a separate coefficient of variation (CV) parameter for each survey, to be estimated. The cv_s parameters are pooled by ages 1, 2-5, 6-9, and 10-15 for the fall and Spanish surveys, and by ages 1, 2, 3-5, 6-9, and 10-15 for the spring survey, as diagnostic model runs suggested differing coefficients of variation at these ages. We treated each survey as from an AR(1) process across ages with independent parameter$s$ ϕ_s to be estimated. A constant CV variance model for I is approximately the same as assuming log(I) has constant variance; however, an advantage of our approach is that we can use observed zero indices directly in the model whereas in other assessment packages these index zeros are typically excluded which is not appropriate when there are many zeros.

Estimation

The unobserved states (i.e. $\delta_{F_{y,a}}, N_{y,a}$) are integrated out and the estimation of all fixed effect parameters (θ) is based on maximizing the marginal likelihood $L(\theta)$,

$$L(\theta) = \iiint_{\Psi} f_\theta(D|\Psi) g_\theta(\Psi) \partial\Psi \tag{1.9}$$

where Ψ is the vector of all random effects, $f_\theta(D|\Psi)$ is the joint probability density function of the data (D) and $g_\theta(\Psi)$is the joint probability density function for the random effects. The TMB (Kristensen *et al.*, 2016) package in R is used to integrate the marginal likelihood (1.8), which is performed via the Laplace approximation (*see* Skaug and Fournier, 2006 for details).

The model fit was assessed by examining the model residuals and retrospective plots. The survey and continuation-ratio logit residual observation model were adjusted to account for observed variance heterogeneity and correlation during model building runs. Retrospective models were fit for years 2005-2017, with each retrospective model fit

using one less year of data (i.e. model for year 2005 used data up to 2005) and predicted abundance, biomass, spawning stock biomass and average *F*'s were plotted and examined for systematic patterns. Ideally, no discernable directional patterns will be present in the retrospective plots.

Biomass-at-age was calculated by multiplying predicted numbers at age (i.e. $N_{y,a}$) and stock weights-at-age, which were estimated externally. A spatiotemporal biphasic Von Bertalanffy growth model (see Kumar *et al.*, 2019) was applied to length-at-age data obtained from the spring research survey data. The combined 3LNO stock weights were estimated by weighting the stock weights for each division by the average index at age during 1975-2017. Stock weights prior to 1975 were fixed at the mean values for 1975-77. Estimated of maturity-at-age were taken from Wheeland *et al.* (2018).

Simulation and sensitivity testing

A full simulation study is beyond the scope of this paper; however, we conducted a simple self-simulation test and jittered start test to examine the reliability of the model estimates (*see e.g.* Cadigan, 2015; Nielsen and Berg, 2014). The self-simulation test randomly generates survey indices and continuation-ratio logit catch proportions from the model predictions and assumed distributions detailed above. Process errors and random effects are treated as fixed and the model is re-fitted to the simulated data. This process is repeated 1000 times and estimates of SSB, average fishing mortality rates (ages 9-14) and recruitment are stored. We calculated the relative difference of the estimates for each year (i.e. (simulation SSB_y – data-based SSB_y)/ data-based SSB_y) for comparison. The jittered start test re-fits the model with random noise added to the starting parameter values, generated from $N(0,0.25 \cdot \hat{\mu})$, where $\hat{\mu}$ is the model predicted parameter of interest. The model is re-optimized 50 times and the negative log-likelihood is stored for each iteration. Ideally, we expect an identical model fit from the jittered starting parameter values. We also examined the model sensitivity to our assumptions about *M*, upper catch bounds and survey catchability, q. We re-fit the model with *M*'s fixed at 0.20 for all ages and years (M2), *M*'s fixed at the original model formulation plus 0.20 (M3), with upper catch bounds fixed at half the original model formulation upper bounds (M4), and with *q* estimated freely at the oldest ages (M5).

Results

The state space model (SSM) fit the data well with no patterns present in the survey or continuation-ratio logit residual plots (omitted for brevity). In 2017 recruitment, abundance and spawning stock biomass (SSB) were

213 estimated near the lowest historical levels (Fig. 1). The model predicted landings were estimated within the upper
214 and lower bounds, with the predicted landings closest to the upper bound in the early 80's, and again in most years
215 since 2006 (Fig. 2) and closest to the lower bound in the early 1990's. At ages 1-4, the catchability pattern (Fig. 3)
216 for the fall and spring surveys was lower for the Engels than the Campelen trawl. The differences were most
217 pronounced for ages three and four, with the catchability estimates for the Campelen trawl almost twice as large as
218 for the Engels trawl. For ages 1-5, the process errors (Fig. 4) were close to zero until the mid-nineties. Overall, there
219 were no noticeable trends in the process errors at the older ages. There were little retrospective patterns (Fig. 5),
220 with SSB slightly overestimated and average fishing mortality for ages 9-14 (aveF) slightly underestimated.
221 In comparison to the most recent stock assessment model for Grand Bank American plaice (which we refer
222 to as the VPA), the overall trends in SSB and aveF were similar (Figs. 6 and 7). Noticeable differences included the
223 SSM predictions of historic SSB (i.e. years 1960-1972) that were larger (but with high uncertainty) than the historic
224 SSB predictions from the VPA. The VPA model also predicted a higher aveF in the early 1990's, at approximately
225 1.1, with the SSM prediction at approximately 0.75 for the same period. From 2005-2009, the estimates for aveF
226 from the SSM were often twice as large as the estimates from the VPA. The retrospective patterns for the SSM were
227 reduced for SSB and greatly reduced for aveF (Fig. 8) compared to the VPA.
228 The self-simulation study lower 10% and upper 90% intervals for both SSB and aveF covered zero (Fig. 9),
229 indicating that the simulated samples produced estimates that were similar to the SSM estimates. In the earliest years
230 (1960-1972), the range of relative differences for aveF was mostly positive, with the converse seen for SSB. Results
231 from the sensitivity tests (Table 2) showed that the SSM had the second lowest BIC overall with M3 (natural
232 mortality rates increased by 0.20) having the lowest BIC. The jittered-start test did not converge for 6% of the
233 simulations, with 77% of the converged models producing negative log-likelihoods that were identical to the
234 original formulation.
235 **Discussion**
236 Overall, our state-space model (SSM) that accounted for uncertainties in the landings data and allowed for
237 process errors fit the data well, with no obvious patterns in the survey and continuation ratio logit residual plots. The
238 retrospective patterns were reduced for spawning stock biomass (SSB) and greatly reduced for average fishing
239 mortality for ages 9-14 (aveF) compared to the most recent stock assessment model (VPA).

The sensitivity runs that reduced M to 0.20 for all ages and years (M2), halved the catch upper bounds (M4), and freely estimated the survey catchabilities (q, M5) all had larger BICs than the SSM. The run that increased the base assumption of M by 0.20 (M3) had a slightly lower BIC than the SSM and this may suggest that the values we used for M's may be too low. Previous research found evidence that M's during 1989 to 1996 (Morgan and Brodie, 2001) had increased to 0.53 and the current VPA model and our SSM include this increase. However, since the closure of the commercial fishery, estimates of total mortality rates have remained high for some periods (*e.g.* Fig. 7 for years 2000-2006), and this may suggest that M is higher than 0.20 in recent years. Preliminary work also suggests that M has increased since the closure of the commercial fishery (Rideout *et al.*, 2008; Morgan *et al.*, 2011). Thus, research that improves our understanding of M for this species should be of high priority.

The SSM retrospective analyses indicated very little systematic patterns, which is a key improvement compared to the VPA model. Including process error in the population dynamics model helped account for underlying time-varying population processes that were not accounted for in the VPA, thereby reducing retrospective patterns. There is still evidence of slight retrospective patterns, and this may be caused by underlying spatial or time-varying process that are mis-specified in the observation model since process errors can only account for mis-specifications in the process equations.

The estimate for survey catchability q is defined as the value required to scale swept-area abundance to the population abundance (*see e.g.* Dickson, 1993;.Fraser, 2007). An estimate of q less than one implies that fewer fish are caught than occupied the area of the trawl, and a value greater than one implies that more fish are caught than occupied the area. Herding behavior of flatfishes in the presence of survey trawls underestimates the width used in area swept calculations and can result in q estimates that are greater than one (Bryan *et al.*, 2013). Therefore, larger q estimates are not unrealistic for American plaice; however, the q estimates from the SSM are very large, with the maximum estimated at 9.1. Thus, additional research is required to better understand why the survey swept area abundance estimates are so much higher than the stock assessment model estimates.

A difference to note between the SSM and the VPA is that the SSM assumes that the survey indices are from a normal distribution with a constant coefficient of variation whereas the VPA assumes that the log of the survey indices are from a lognormal distribution. The lognormal distribution does not allow for zeros in the survey data, however this assumption may not be appropriate when there are many zeros in the data or when zeros are

"true" zeros (i.e. no fish available to be caught). The assumption of normality with a constant coefficient of variation avoids the problem of dropping zeros altogether. Although the methods are theoretically similar, future research is needed to compare the performance of the two.

Our new SSM is an improvement to the VPA current stock assessment model that is used to inform the management of American plaice on the Grand Bank of Newfoundland as it allows for errors in the landings data and reduces the retrospective patterns. This model is also a valuable first step in improving our understanding of the stock as the flexibility of state-space models are an ideal foundation from which to build more complex and realistic models.

Acknowledgements

Research funding was provided by the Ocean Frontier Institute, through an award from the Canada First Research Excellence Fund. Research funding to NC was also provided by the Ocean Choice International Industry Research Chair program at the Marine Institute of Memorial University of Newfoundland. Funding to AP was also provided by a Natural Sciences and Engineering Research Council of Canada Master's Graduate Scholarship. Many thanks are also extended to Dr. Anders Nielsen, Danish Technical University, for advice on more computationally efficient ways to implement our model in TMB.

References

Aeberhard, W.H., J. Mills-Flemming, A. Nielsen. 2018. A review of state-space models for fisheries science. Annual Review of Statistics and its Application, 5: 215-23. http://dx.doi.org/10.1146/annurev-statistics-031017-100427

Agresti, A. 2003. Categorical data analysis. Vol. 482. John Wiley & Sons. http://dx.doi.org/10.1002/0471249688

Albertsen, C.M., A. Nielsen and U.H. Thygesen. 2016. Choosing the observational likelihood in state-space stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences, 74(5):779-789. http://dx.doi.org/10.1139/cjfas-2015-0532

Berg, C. W. and K. Kristensen. 2012. Spatial age-length key modelling using continuation ratio logits. Fisheries Research, 129: 119–126. http://dx.doi.org/10.1016/j.fishres.2012.06.016

Brodie, W.B. MS 1986. An assessment of the American plaice stock on the Grand Bank (NAFO Divisions 3LNO). NAFO SCR Doc., No. 86, Serial No, 41.

Bryan, D. R., K.L. Bosley, A.C. Hicks, M.A. Haltuch and W.W. Wakefield. 2014. Quantitative video analysis of flatfish herding behavior and impact on effective area swept of a survey trawl. Fisheries Research, 154: 120-126. https://doi.org/10.1016/j.fishres.2014.02.007

Cadigan, N. G. 2015. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. Canadian Journal of Fisheries and Aquatic Sciences, 73(2): 296-308. http://dx.doi.org/10.1139/cjfas-2015-0047

DFO. MS 2011. Recovery potential assessment of American plaice (*Hippoglossoides platessoides*) in Newfoundland and Labrador. DFO Can. Sci. Advis. Sec., Sci. Advis. Rep. 2011/030.

Dickson, W. 1993. Estimation of the capture efficiency of trawl gear. I: development of a theoretical model. Fisheries Research, 16(3): 239-253. http://dx.doi.org/10.1016/0165-7836(93)90096-P

Dwyer K., R. Rideout, D. Ings, D. Power, M. Morgan, B. Brodie and P.B. Healy. MS 2016. Assessment of American plaice in Div. 3LNO. NAFO SCS Doc. No. 16, Serial No. 030.

Dwyer K., M. Morgan, B. Brodie, D. Maddock Parsons, R. Rideout, P.B. Healy and D. Ings. MS 2014. Survey indices and STATLANT 21A bycatch information for American plaice in NAFO Div. 3LNO. NAFO SCR Doc. No. 14, Serial No. 031.

Fournier, D., and C.P. Archibald. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences, 39(8): 1195-1207. https://doi.org/10.1139/f82-157

Fraser, H.M., S.P. Greenstreet, and G.J. Piet. 2007. Taking account of catchability in groundfish survey trawls: implications for estimating demersal fish biomass. ICES Journal of Marine Science, 64:(9). 1800-1819. http://dx.doi.org/10.1093/icesjms/fsm145

González-Troncoso, D., A. Gago1, A. Nogueira and E. Román. MS 2017. Results for Greenland halibut, American plaice and Atlantic cod of the Spanish survey in NAFO Div. 3NO for the period 1997-2016. NAFO SCR Doc. No. 17, Serial No. 018.

Hilborn, R. and C.J. Walters. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Reviews in Fish Biology and Fisheries, 2(2): 177-178. http://dx.doi.org/10.1007/978-1-4615-3598-0

Kristensen, K., A. Nielsen, C.W. Berg, H. Skaug and B.M. Bell. 2016. TMB: Automatic differentiation and Laplace approximation. Journal of Statistical Software, 70: 1–26. http://dx.doi.org/10.18637/jss.v070.i05

Kumar, R., N.G. Cadigan, N. Zheng, D. Varkey, and M.J. Morgan. 2019. A spatial and state-space survey-based assessment model applied to the Grand Bank, Newfoundland and Labrador stock of American plaice. Submitted to journal.

Legault C.M., Chair. MS 2009. Report of the Retrospective Working Group, January 14-16, 2008, Woods Hole, Massachusetts. Northeast Fish Sci Cent Ref Doc. 09-01 30.

Maunder, M.N., and A.E. Punt. 2013. A review of integrated analysis in fisheries stock assessment. Fisheries Research, 142: 61-74. https://doi.org/10.1016/j.fishres.2012.07.025

Megrey, B.A. MS 1989. Review and comparison of age-structured stock assessment models. *In Amer. Fish. Symp.* 6: 8-48.

Mohn, R. 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES Journal of Marine Science, 56(4): 473-488. http://dx.doi.org/10.1006/jmsc.1999.0481

Morgan, M.J., and W.B. Brodie. MS 2001. An exploration of virtual population analyses for Divisions 3LNO American plaice. NAFO SCR Doc., No. 1, Serial No. 4.

Morgan, M.J., J. Bailey, B.P. Healey, D. Maddock Parsons, and R. Rideout. MS 2011. Recovery potential assessment of American Plaice (Hippoglossoides platessoides) in Newfoundland and Labrador. DFO Can. Sci. Advis. Sec. Res. Doc. 2011/047.

NAFO. MS 2017. Report of the NAFO Joint Commission-Scientific Council Working Group on Catch Reporting and NAFO Ad hoc Joint Commission-Scientific Council Catch Data Advisory Group (CDAG) Meeting. NAFO COM-SC. 17-08.

Nielsen, A., and C.W. Berg. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research, 158: 96-101. http://dx.doi.org/10.1016/j.fishres.2014.01.014

Pitt, T.K. MS 1972. Nominal catches of American plaice in Divisions 3L and 3N for the years 1960-1970. ICNAF Res. Doc, 72/90.

Rideout, R.M., M.J. Morgan and C. Busby. MS 2008. COSEWIC Status Report on American plaice Hippoglossoides platessoides..

Rivard, D. MS 1980. Back-calculating production from cohort analysis, with discussion on surplus production for two redfish stocks. CAFSAC Res. Doc., 80/23.

Shelton, P.A. and M.J. Morgan. 2005. Is by-catch mortality preventing the rebuilding of cod (*Gadus morhua*) and American plaice (*Hippoglossoides platessoides*) stocks on the Grand Bank. Journal of Northwest Atlantic Fisheries Science. 36: 1-17. http://dx.doi.org/10.2960/J.v36.m544

Skaug, H.J. and D.A. Fournier. 2006. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. Computational Statistics & Data Analysis, 51(2): 699-709. http://dx.doi.org/10.1016/j.csda.2006.03.005

Szuwalski, C.S., J.N. Ianelli and A.E. Punt. 2017. Reducing retrospective patterns in stock assessment and impacts on management performance. ICES Journal of Marine Science, 75(2): 596-609. http://dx.doi.org/10.1093/icesjms/fsx159

Wheeland, L., K. Dwyer, M. Morgan, R. Rideout and R. Rogers. MS 2018. Assessment of American plaice in Div. 3LNO. NAFO SCS Doc. No. 18, Serial No. 039.

Zheng, N. and N. Cadigan. 2019. Likelihood methods for basic stratified sampling, with application to Von Bertalanffy growth model estimation. Open Journal of Statistics, 9: 623-642. https://doi.org/10.4236/ojs.2019.96040

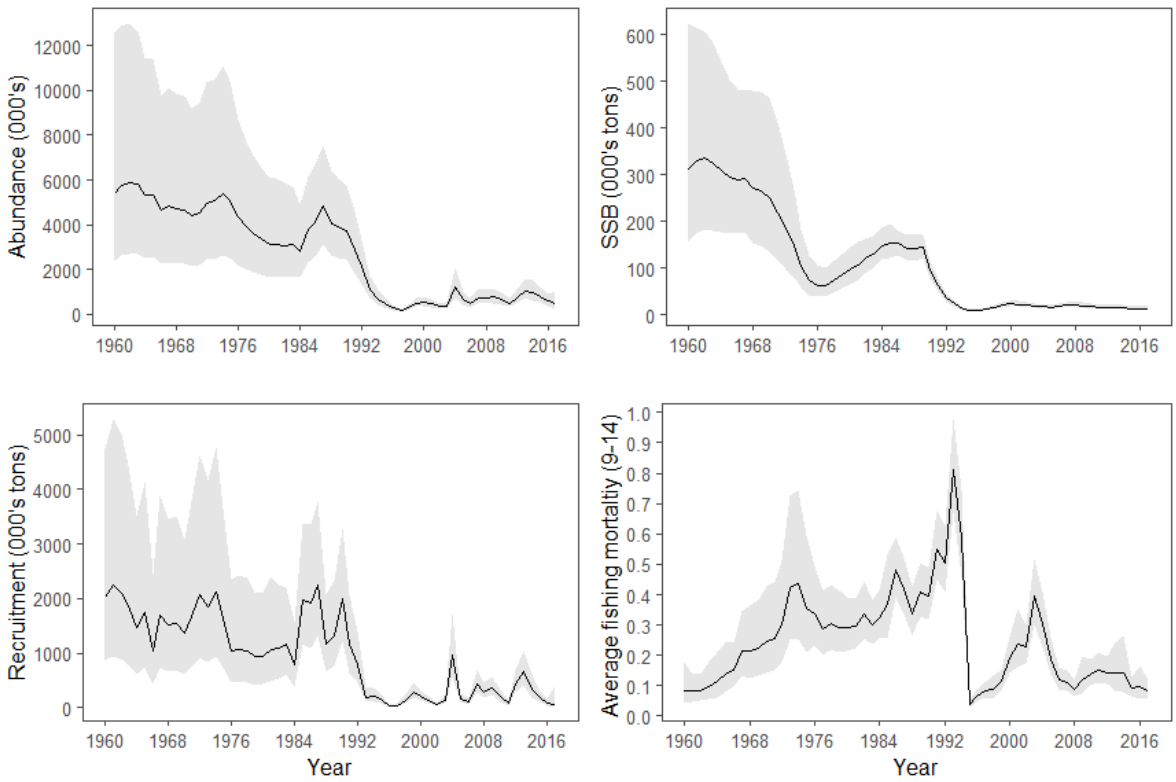
Table 1. Upper catch bounds (UB) for estimated landings with associated justification for bounds; RC is reported catch. Discussion on catch uncertainties can be found in Wheeland et al. 2018, and references therein

Period	UB	Comments
1960-1976	2xRC	"Unspecified flounder" by some countries, see, for example, (Pitt, 1972).
1977-1982	1.2xRC	Landings by primarily Canada (>95%) after establishment of 200 mile EEZ
1983-1993	1.5xRC	increased foreign catch outside 200 miles; Various estimates used for catch; Issues with unspecified flounder records and discarding
1994-2010	1.2xRC	No directed fishing in 1994 (bycatch quota), 0 TAC 1995 onwards; Catches defined from various sources with those considered most reliable by Scientific Council used for totals
2011-2017	1.5xRC	Loss of availability of scientific observer data in the NAFO Regulatory area and surveillance estimates. Varying methods applies to obtain catch estimates including effort ratios (Dwyer *et al.*, 2016), daily catch records, and NAFO CESAG estimates (NAFO, 2017)

Table 2. Model estimates of variance parameters and some population parameters with percent coefficient of variation. (M1 is base model, M2 is base model with natural mortality=0.20, M3 is base model with natural mortality increased by 0.20, M4 is base model with upper catch bounds set to half the base model upper bounds, M5 is base model with catchability freely estimated)

	SSM		**M2**		**M3**		**M4**		**M5**	
	nll = 4516 AIC=9192 BIC=9595		nll = 4534 AIC=9228 BIC=9631		nll = 4489 AIC=9137 BIC=9541		nll = 4546 AIC=9253 BIC=9656		nll = 4479 AIC=9148 BIC=9627	
	Est	**CV**	**Est**	**CV**	**Est**	**CV**	**Est**	**CV**	**Est**	**CV**
σ_{F5}	1.09	40	1.37	39	0.97	34	1.09	38	0.43	62
σ_{F6+}	0.81	22	0.86	20	0.94	24	0.83	22	0.52	39
σ_{pe}	0.21	6	0.22	6	0.20	6	0.21	6	0.20	6
σ_{R}	0.70	14	0.98	14	0.65	13	0.67	14	0.64	13
$\text{cv}_{\text{Fall 1}}$	0.69	20	0.68	18	0.71	20	0.69	20	0.71	19
$\text{cv}_{\text{Fall 2-5}}$	0.33	10	0.33	10	0.34	10	0.33	10	0.34	10
$\text{cv}_{\text{Fall 6-9}}$	0.25	10	0.26	10	0.25	10	0.25	10	0.26	10
$\text{cv}_{\text{Fall 10-15}}$	0.43	9	0.44	9	0.42	9	0.43	9	0.42	9
$\text{cv}_{\text{Span. 1}}$	2.12	47	2.05	46	2.18	48	2.12	47	2.09	46
$\text{cv}_{\text{Span. 2-5}}$	1.05	22	1.03	21	1.04	21	1.04	21	1.03	21
$\text{cv}_{\text{Span. 6-9}}$	0.55	14	0.54	14	0.54	14	0.54	14	0.53	14
$\text{cv}_{\text{Span. 10-15}}$	0.47	12	0.45	11	0.46	12	0.46	12	0.46	12
$\text{cv}_{\text{Spr. 1}}$	1.60	33	1.55	32	1.48	31	1.61	33	1.46	30
$\text{cv}_{\text{Spr. 2}}$	0.82	21	0.80	20	0.75	19	0.82	21	0.75	19
$\text{cv}_{\text{Spr. 3-5}}$	0.47	15	0.45	15	0.44	14	0.48	15	0.44	14
$\text{cv}_{\text{Spr. 6-9}}$	0.29	12	0.28	12	0.27	11	0.29	12	0.27	11
$\text{cv}_{\text{Spr. 10-15}}$	0.38	11	0.36	11	0.35	10	0.38	11	0.35	11
$\varphi_{F_{A6+}}$	0.99	1	0.99	1	0.99	1	0.99	1	1.00	2
φ_{F5}	0.95	5	0.96	4	0.91	7	0.94	6	0.74	41
$\varphi_{F_{Y6+}}$	0.94	3	0.95	2	0.94	4	0.92	4	0.87	11
φ_{R}	0.33	45	0.66	16	0.31	47	0.30	49	0.28	52
φ_{Fall}	0.57	9	0.59	9	0.58	9	0.56	9	0.59	9
$\varphi_{\text{Span.}}$	0.72	6	0.71	6	0.71	6	0.71	6	0.72	6
$\varphi_{\text{Spr.}}$	0.88	3	0.87	3	0.86	3	0.88	3	0.86	3
φ_{C_Y}	0.25	29	0.28	27	0.24	30	0.24	31	0.43	23
φ_{C_A}	0.81	4	0.81	4	0.80	4	0.80	4	0.86	3
F_{2017}	0.08	19	0.12	16	0.02	20	0.07	18	0.02	26
SSB_{2017}	13.63	18	9.59	15	48.01	18	13.88	17	49.19	25
F_{VPA}	0.06	-	-	-	-	-	-	-	-	-
SSB_{VPA}	18.24	-	-	-	-	-	-	-	-	-

25 Figure 1.



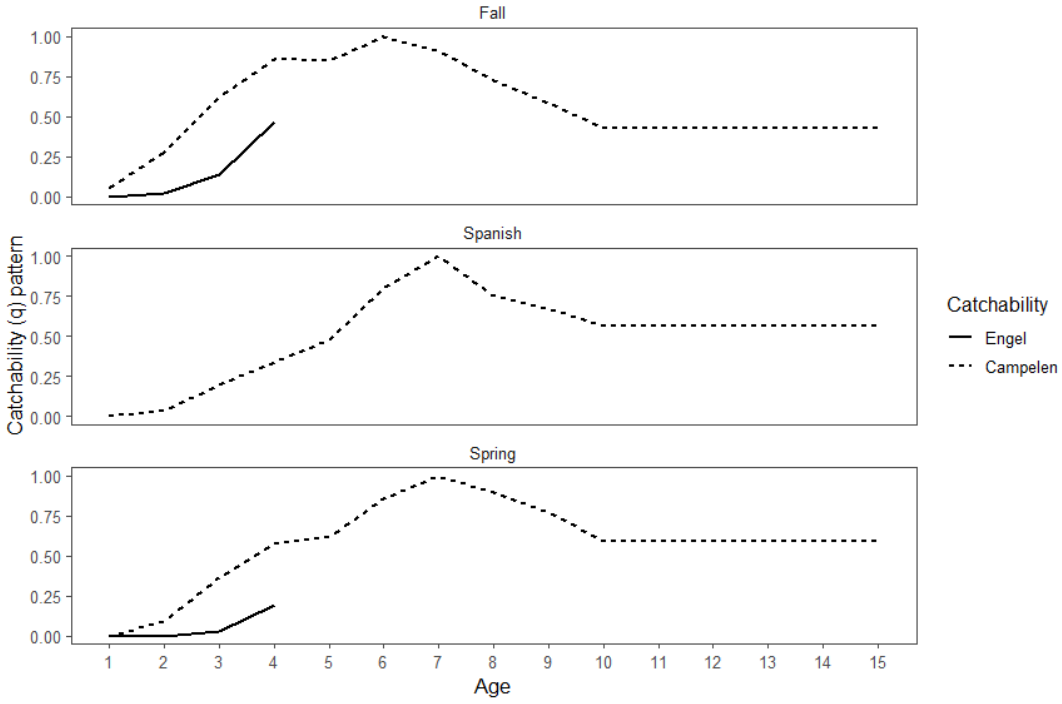
26

27 Figure 2.

28

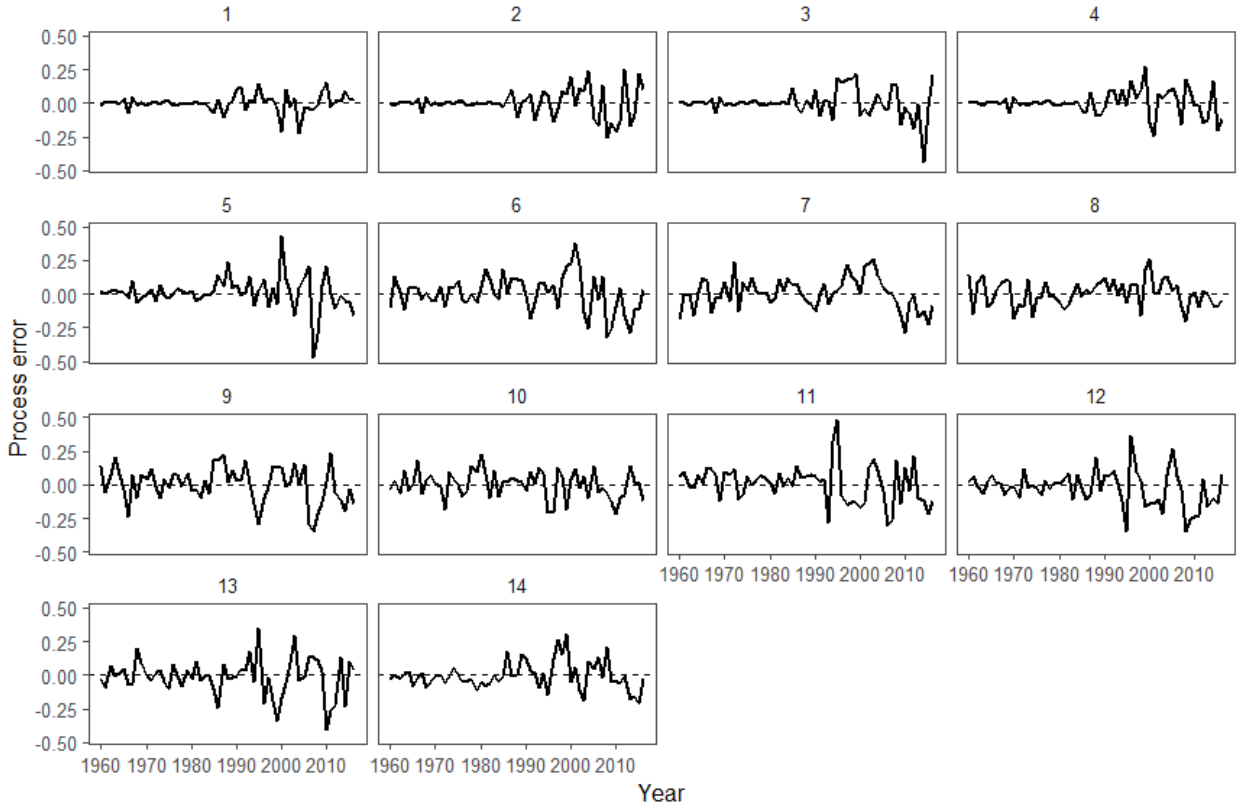
29

30 Figure 3.



31

32 Figure 4.



33

Figure 5.

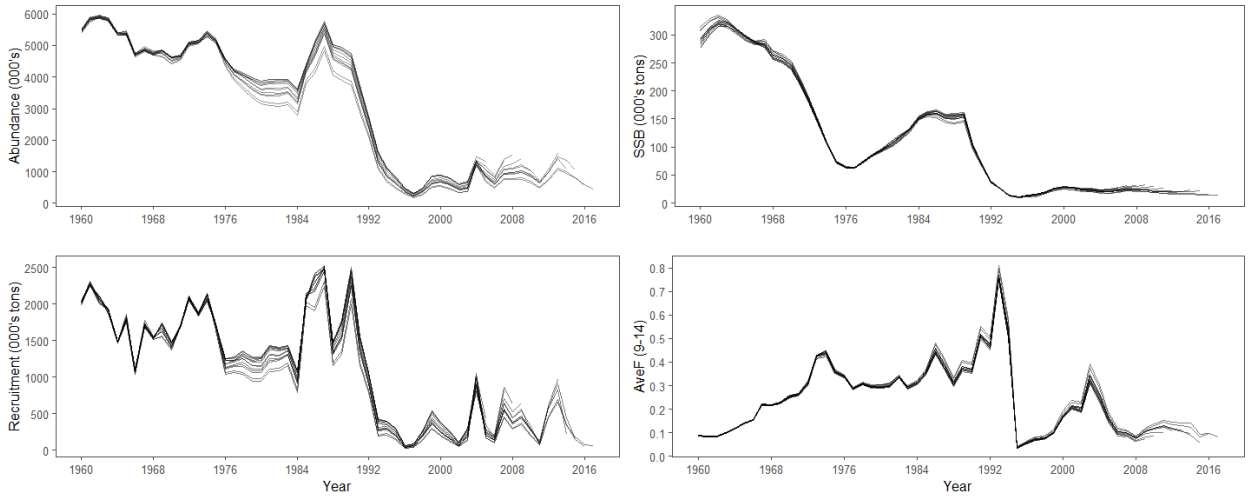


Figure 6.

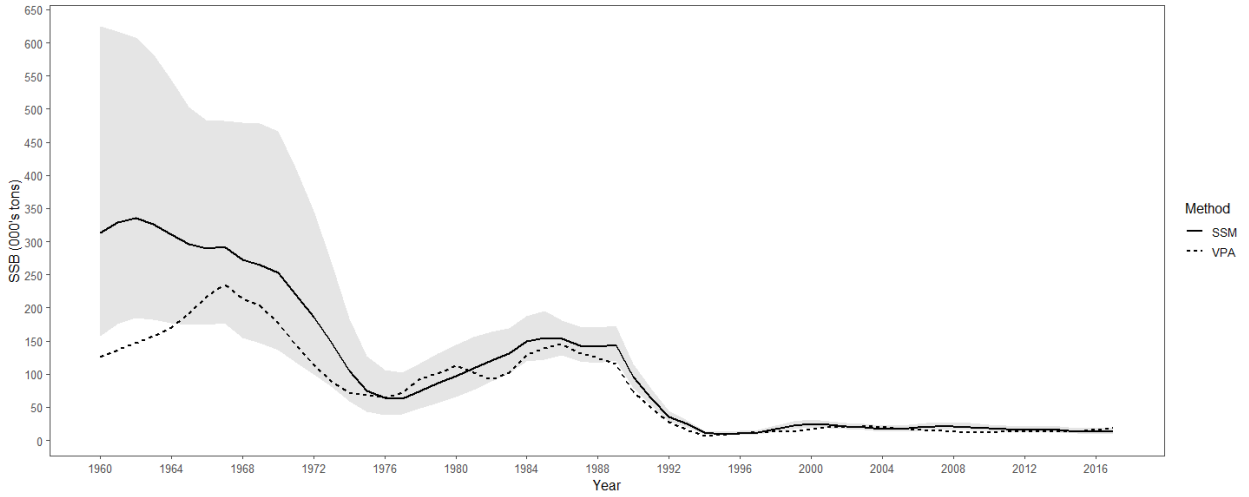


Figure 7.

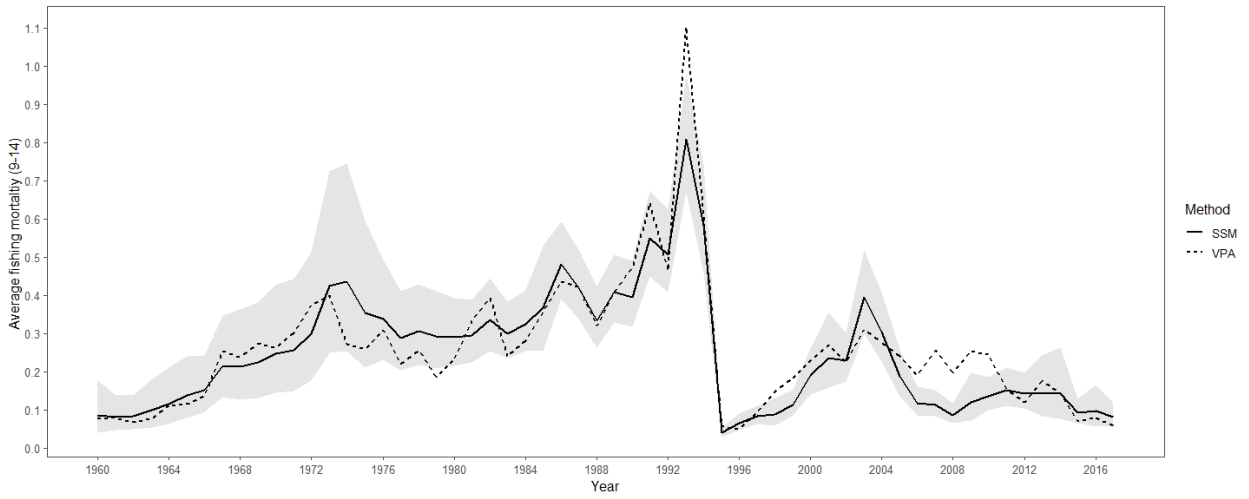


Figure 8.

Figure 9.

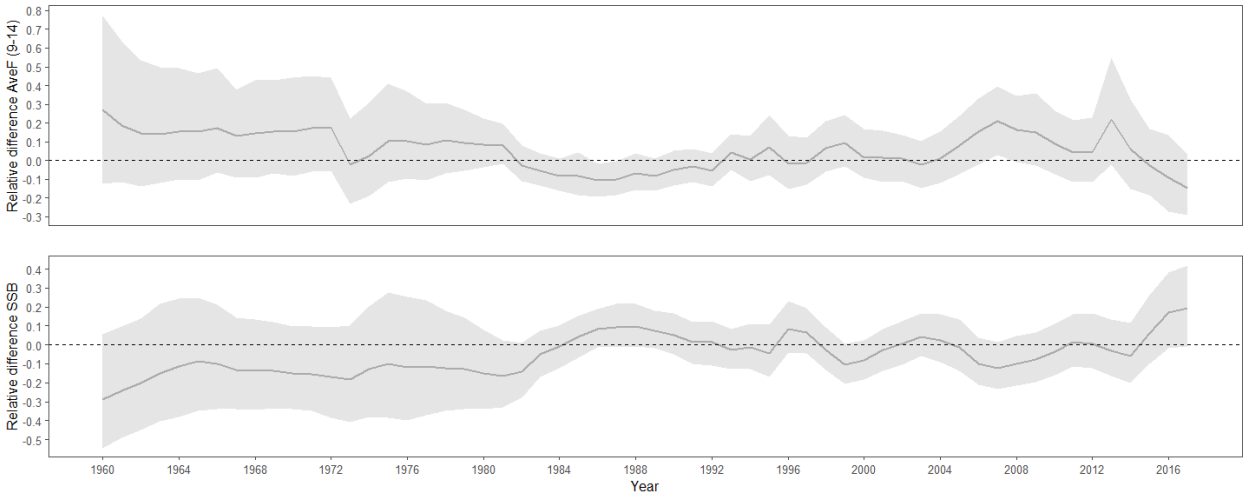

Relative difference AveF (9-14)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
Relative difference SSB
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
1960
1964
1968
1972
1976
1980
1984
1988
1992
1996
2000
2004
2008
2012
2016
Year

Figure 1. Model estimated population abundance, spawning stock biomass (SSB), average fishing mortality rate (ages 9-14) and recruitment.

Figure 2. Model estimated log catch numbers for ages 5-15+ (solid line), the shaded grey represents the region between the log lower catch bounds and the log upper catch bounds.

Figure 3. Survey catchability patterns for the fall, spring and Spanish surveys. A separate catchability parameter is estimated for two gear types (Engel and Campelen) for the spring and fall surveys for ages 1-4.

Figure 4. Model estimated process errors for ages 1-14 for years 1960-2016.

Figure 5. Retrospective estimates from 2008-2017 for total abundance, spawning stock biomass, average fishing mortality rates (ages 9-14) and recruitment for years 1960-2017.

Figure 6. Model (SSM) and VPA estimated spawning stock biomass for years 1960-2017.

Figure 7. Model (SSM) and VPA estimated average fishing mortality rates (ages 9-14) for years 1960-2017.

Figure 8. Model (SSM) and VPA retrospective estimates from 2013-2017 of average fishing mortality rates (ages 9-14) and of spawning stock biomass for years 2014-2017.

Figure 9. Relative difference from model (SSM) and self-simulated sample for spawning stock biomass and average fishing mortality rates (ages 9-14). The solid grey line in the median of the estimates and shaded grey regions represent the lower 10% and upper 90% bounds.

Bibliography

Aeberhard, W. H., Mills Flemming, J. & Nielsen, A. (2018), 'Review of state-space models for fisheries science', Annual Review of Statistics and Its Application **5**, 215–235.

Agger, P., Boetius, I. & Lassen, H. (1971), 'On errors in the virtual population analysis', International Council of the Exploration of the Sea, CM .

Ailloud, L. E. & Hoenig, J. M. (2019), 'A general theory of age-length keys: combining the forward and inverse keys to estimate age composition from incomplete data', ICES Journal of Marine Science **76**(6), 1515–1523.

Albertsen, C. M., Nielsen, A. & Thygesen, U. H. (2017), 'Connecting single-stock assessment models through correlated survival', ICES Journal of Marine Science **75**(1), 235–244.

Ballesteros, M., Chapela, R., Ramírez-Monsalve, P., Raakjaer, J., Hegland, T. J., Nielsen, K. N., Laksá, U. & Degnbol, P. (2017), 'Do not shoot the messenger: Ices advice for an ecosystem approach to fisheries management in the european union', ICES Journal of Marine Science **75**(2), 519–530.

Baranov, T. (1918), 'On the question of the biological basis of fisheries. nauchnyi issledovatelskii ikhtiologicheskii institut isvestia 1 (1): 81–128', Reports from the Division of Fish Management and Scientific Study of the Fishing Industry.](English translation by WE Ricker, 1945. Mimeographed.) .

Barbeaux, S., J. I. & Paulson, W. (2018), 'Aleutian islands walleye pollock safe: Stock assessment and fishery evaluation report for the groundfish resources of the bering sea/aleutian islands regions', North Pac. Fish. Mgmt. Council .

Beare, D., Needle, C., Burns, F. & Reid, D. (2005), 'Using survey data independently from commercial data in stock assessment: an example using haddock in ices division via', ICES Journal of Marine Science **62**(5), 996–1005.

Brodie, W. (1986), 'An assessment of the american plaice stock on the grand bank (nafo divisions 3lno).', NAFO SCR Doc **86**, 41.

Brooks, L. (2017), 'Georges bank haddock', Operational Assessment of 19 Northeast Groundfish Stocks, Updated Through 2016 .

Bull, B., Francis, R., Dunn, A., McKenzie, A., Gilbert, D., Smith, M., Bian, R. & Fu, D. (2005), 'Casal (c++ algorithmic stock assessment laboratory): Casal user manual v2'.

Cadigan, N. G. (2015), 'A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates', Canadian Journal of Fisheries and Aquatic Sciences **73**(2), 296–308.

Chen, J., Thompson, M. E. & Wu, C. (2004), 'Estimation of fish abundance indices based on scientific research trawl surveys', Biometrics **60**(1), 116–123.

Chouinard, G. & Millar, C. (2019), 'Retrospective patterns in age-structured fish stock assessments: Mohn's rho estimates from the 2018 and 2019 ices assessments', Presentation.

Christensen, V. & Walters, C. (2000), 'Ecopath with ecosim: Methods, capabilities and limitations.', Fish. Cent. Res. Rep. **8**(2), 79–105.

Ciannelli, L., Fisher, J. A., Skern-Mauritzen, M., Hunsicker, M. E., Hidalgo, M., Frank, K. T. & Bailey, K. M. (2013), 'Theory, consequences and evidence of eroding population spatial structure in harvested marine fishes: a review', Marine Ecology Progress Series **480**, 227–243.

Clark, W. (1981), 'Restricted least-squares estimates of age composition from length composition', Canadian Journal of Fisheries and Aquatic Sciences **38**(3), 297–307.

Cook, R. M., Holmes, S. J. & Fryer, R. J. (2015), 'Grey seal predation impairs recovery of an over-exploited fish stock', Journal of Applied Ecology **52**(4), 969–979.

Cotter, A., Burt, L., Paxton, C., Fernandez, C., Buckland, S. & Pan, J.-X. (2004), 'Are stock assessment methods too complicated?', Fish and Fisheries **5**(3), 235–254.

Cotter, A. & Pilling, G. (2007), 'Landings, logbooks and observer surveys: improving the protocols for sampling commercial fisheries', Fish and Fisheries **8**(2), 123–152.

Darby, C. & Flatman, S. (1994), Virtual population analysis: version 3.1 (Windows/Dos) user guide, Great Britain, Ministry of Agriculture, Fisheries and Food, Directorate of ....

Deriso, R., Quinn Ii, T. & Neal, P. (1985), 'Catch-age analysis with auxiliary information', Canadian Journal of Fisheries and Aquatic Sciences **42**(4), 815–824.

DFO (2017), 'Assessment of the status of division 4x5y haddock in 2009', DFO Can. Sci. Advis. Sec. Sci. Advis. Rep **2017/006**.

Dichmont, C. M., Deng, R. A., Punt, A. E., Brodziak, J., Chang, Y.-J., Cope, J. M., Ianelli, J. N., Legault, C. M., Methot Jr, R. D., Porch, C. E. et al. (2016), 'A review of stock assessment packages in the united states', Fisheries Research **183**, 447–460.

Dolan, T. E., Patrick, W. S. & Link, J. S. (2015), 'Delineating the continuum of marine ecosystem-based management: a us fisheries reference point perspective', ICES Journal of Marine Science **73**(4), 1042–1050.

Doubleday, W. (1976), 'A least squares approach to analyzing catch at age data', Int. Comm. Northwest Atl. Fish. Res. Bull **12**(1), 69–81.

FNZ (2019), 'Fisheries assessment plenary, may 2019: stock assessments and stock status.', New Zealand Ministry of Fisheries, Wellington .

Fournier, D. A., Hampton, J. & Sibert, J. R. (1998), 'Multifan-cl: a length-based, age-structured model for fisheries stock assessment, with application to south pacific albacore, thunnus alalunga', Canadian Journal of Fisheries and Aquatic Sciences **55**(9), 2105–2116.

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A. & Sibert, J. (2012), 'Ad model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models', Optimization Methods and Software **27**(2), 233–249.

Fournier, D. & Archibald, C. P. (1982), 'A general theory for analyzing catch at age data', Canadian Journal of Fisheries and Aquatic Sciences **39**(8), 1195–1207.

Fournier, D., Sibert, J. R., Majkowski, J. & Hampton, J. (1990), 'Multifan a likelihood-based method for estimating growth parameters and age composition from multiple length frequency data sets illustrated using data for southern bluefin tuna (thunnus maccoyii)', Canadian journal of fisheries and aquatic sciences **47**(2), 301–317.

Francis, R. C. (2011), 'Data weighting in statistical fisheries stock assessment models', Canadian Journal of Fisheries and Aquatic Sciences **68**(6), 1124–1138.

Francis, R. C. (2014), 'Replacing the multinomial in stock assessment models: A first step', Fisheries Research **151**, 70–84.

Francis, R. C. (2017), 'Revisiting data weighting in fisheries stock assessment models', Fisheries Research **192**, 5–15.

Fridriksson, A. (1934), 'On the calculation of age-distribution within a stock of cod by means of relatively few age-determinations as a key to measurements on a large scale', Rapports Et Proces-Verbaux Des Reunions, Conseil International Pour l'Exploration De La Mer **86**, 1–5.

Fry, F. E. (1949), 'Statistics of a lake trout fishery', Biometrics **5**(1), 27–67.

Gavaris, S. (1988), 'An adaptive framework for the estimation of population size', Res. Doc. Can. Atl. Fish. Scient. Adv. Comm. **88**, 1–12.

Gavaris, S. & Gavaris, C. (1983), 'Estimation of catch at age and its variance for groundfish stocks in the newfoundland region', Sampling commercial catches of marine fish and invertebrates. Edited by WG Doubleday and D. Rivard. Can. Spec. Publ. Fish. Aquat. Sci **66**, 178–182.

Gudmundsson, G. (1994), 'Time series analysis of catch-at-age observations', Journal of the Royal Statistical Society: Series C (Applied Statistics) **43**(1), 117–126.

Gulland, J. (1965), 'Estimation of mortality rates. annex to arctic fisheries working group report', International Council for the Exploration of the Sea, CM .

Haddon, M. (2010), Modelling and quantitative methods in fisheries, CRC press.

Hampton, J. & Fournier, D. A. (2001), 'A spatially disaggregated, length-based, age-structured population model of yellowfin tuna (thunnus albacares) in the western and central pacific ocean', Marine and Freshwater Research **52**(7), 937–963.

Harvey, C. J., Kelble, C. R. & Schwing, F. B. (2016), 'Implementing "the iea": using integrated ecosystem assessment frameworks, programs, and applications in support of operationalizing ecosystem-based management', ICES Journal of Marine Science **74**(1), 398–405.

Helser, T. E., Punt, A. E. & Methot, R. D. (2004), 'A generalized linear mixed model analysis of a multi-vessel fishery resource survey', Fisheries Research **70**(2-3), 251–264.

Hilborn, R. & Walters, C. J. (2013), Quantitative fisheries stock assessment: choice, dynamics and uncertainty, Springer Science & Business Media.

Hoenig, J. M., Choudary Hanumara, R. & Heisey, D. M. (2002), 'Generalizing double and triple sampling for repeated surveys and partial verification', Biometrical Journal: Journal of Mathematical Methods in Biosciences **44**(5), 603–618.

Holmes, S. & Fryer, R. (2011), 'Significance of seal feeding on cod west of scotland–results from a state space stock assessment model', ICES CM **1**, 22.

ICES (2010), 'Report of the workshop on reviews of recent advances in stock assessment models worldwide: "around the world in ad models" (wkadsam), 27 september - 1 october 2010, nantes, france', CM **2010/SSGSUE**(10), 120.

ICES (2011), 'Report of the workshop on the analysis of the benchmark of cod in subarea iv (north sea), division viid (eastern channel) and division iiia (skagerrak)(wkcod 2011), 7–9 february 2011, copenhagen, denmark'.

ICES (2019a), 'Cod (gadus morhua) in divisions 7.e–k (western english channel and southern celtic seas)', Report of the ICES Advisory Committee **cod.27.7e–k**.

ICES (2019b), 'Sole (solea solea) in division 7.d (eastern english channel)', Report of the ICES Advisory Committee **sol.27.7d**.

ICES (2019c), 'Sole (solea solea) in subdivisions 20–24 (skagerrak and kattegat, western baltic sea)', Report of the ICES Advisory Committee, 2019. .

ISC, B. W. G. (2018a), 'Stock assessment for swordfish (xiphias gladius) in thewestern and central north pacific ocean through 2016', REPORT OF THE BILLFISH WORKING GROUP .

ISC, S. W. G. (2017), 'Stock assessment and future projections of blue shark in the north pacific ocean through 2015', REPORT OF THE SHARK WORKING GROUP .

ISC, S. W. G. (2018b), 'Stock assessment of shortfin mako shark in the north pacific ocean through 2016', REPORT OF THE SHARK WORKING GROUP .

Jones, R. (1961), The assessment of the long term effects of changes in gear selectivity and fishing effort, HM Stationery Office.

Kimura, D. K. & Chikuni, S. (1987), 'Mixtures of empirical distributions: an iterative application of the age-length key', Biometrics pp. 23–35.

Knape, J. & De Valpine, P. (2012), 'Fitting complex population models by combining particle filters with markov chain monte carlo', Ecology **93**(2), 256–263.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. & Bell, B. (2015), 'Tmb: automatic differentiation and laplace approximation', arXiv preprint arXiv:1509.00660 .

Lassen, H. & Medley, P. (2001), Virtual population analysis: a practical manual for stock assessment, number 400, Food & Agriculture Org.

Legault, C. M. & Restrepo, V. R. (1998), 'A flexible forward age-structured assessment program', ICCAT. Col. Vol. Sci. Pap **49**, 246–253.

Lowe, S., J. I. & Paulson, W. (2018), 'Stock assessment of aleutian islands atka mackerel: Stock assessment and fishery evaluation report for the groundfish resources of the bering sea/aleutian islands regions', North Pac. Fish. Mgmt. Council .

MacCall, A. D. (2009), 'Depletion-corrected average catch: a simple formula for estimating sustainable yields in data-poor situations', ICES Journal of marine Science **66**(10), 2267–2271.

Maunder, M. N. & Piner, K. R. (2014), 'Contemporary fisheries stock assessment: many issues still remain', ICES Journal of Marine Science **72**(1), 7–18.

Maunder, M. N., Schnute, J. T. & Ianelli, J. N. (2009), Computers in fisheries population dynamics, in 'Computers in fisheries research', Springer, pp. 337–372.

Megrey, B. A. (1989), Review and comparison of age-structured stock assessment models, in 'Amer. Fish. Symp.', Vol. 6, pp. 8–48.

Methot Jr, R. D. & Wetzel, C. R. (2013), 'Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management', Fisheries Research **142**, 86–99.

Morgan, M. J., Bailey, J., Healey, B., Maddock Parsons, D. & Rideout, R. (2011), Recovery potential assessment of American Plaice (Hippoglossoides platessoides) in Newfoundland and Labrador, Fisheries and Oceans Canada, Science.

Murphy, G. I. (1965), 'A solution of the catch equation', Journal of the Fisheries Board of Canada **22**(1), 191–202.

Naumann, U. & Schenk, O. (2012), Getting started with adol-c, in 'Combinatorial Scientific Computing', Chapman and Hall/CRC, pp. 205–226.

NEFSC (2018), '64th northeast regional stock assessment workshop (64th saw) assessment summary report', US Dept Commer Northeast Fish Sci Cent Ref Doc **18-03**.

Nielsen, A. & Berg, C. W. (2014), 'Estimation of time-varying selectivity in stock assessments using state-space models', Fisheries Research **158**, 96–101.

Ogle, D. H. (2016), Introductory fisheries analyses with R, Chapman and Hall/CRC.

Palmer, M. (2017), 'Gulf of maine atlantic cod', Operational Assessment of 19 Northeast Groundfish Stocks, Updated Through 2016 .

Paloheimo, J. (1980), 'Estimation of mortality rates in fish populations', Transactions of the American Fisheries Society **109**(4), 378–386.

Pennington, M. & Volstad, J. H. (1994), 'Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys', Biometrics pp. 725–732.

Pitt, T. (1972), 'Nominal catches of american plaice in divisions 3l and 3n for the years 1960-1970', International Commission for the Norwest Atlantic Fisheries **72**(90).

Plagányi, É. E., Punt, A. E., Hillary, R., Morello, E. B., Thébaud, O., Hutton, T., Pillans, R. D., Thorson, J. T., Fulton, E. A., Smith, A. D. et al. (2014), 'Multispecies fisheries management and conservation: tactical applications using models of intermediate complexity', Fish and Fisheries **15**(1), 1–22.

Pope, J. (1972), 'An investigation of the accuracy of virtual population analysis using cohort analysis', ICNAF Research Bulletin **9**(10), 65–74.

Pope, J. & Shepherd, J. (1982), 'A simple method for the consistent interpretation of catch-at-age data', ICES Journal of Marine Science **40**(2), 176–184.

Pope, J. & Shepherd, J. (1985), 'A comparison of the performance of various methods for tuning vpas using effort data', Ices Journal of Marine Science **42**(2), 129–151.

Punt, A., Dunn, A., Elvarsson, B., Hampton, J., Hoyle, S., Maunder, M., Methot, R. & Nielsen, A. (2020), 'Essential features of the next-generation integrated fisheries stock assessment package: A perspective', Fisheries Research **229**, 111–115.

Quinn, T. J. (2003), 'Ruminations on the development and future of population dynamics models in fisheries', Natural Resource Modeling **16**(4), 341–392.

R. Rideout, D. I. & Brattey, J. (2017), 'An assessment of the cod stock in nafo divisions 3no', NAFO SCR Doc **17**, 042.

Shelton, A. O., Thorson, J. T., Ward, E. J. & Feist, B. E. (2014), 'Spatial semiparametric models improve estimates of species abundance and distribution', Canadian Journal of Fisheries and Aquatic Sciences **71**(11), 1655–1666.

Shepherd, J. (1999), 'Extended survivors analysis: An improved method for the analysis of catch-at-age data and abundance indices', ICES Journal of Marine Science **56**(5), 584–591.

Skern-Mauritzen, M., Ottersen, G., Handegard, N. O., Huse, G., Dingsør, G. E., Stenseth, N. C. & Kjesbu, O. S. (2016), 'Ecosystem processes are rarely included in tactical fisheries management', Fish and Fisheries **17**(1), 165–175.

Smith, S. J. (1990), 'Use of statistical models for the estimation of abundance from groundfish trawl survey data', Canadian Journal of Fisheries and Aquatic Sciences **47**(5), 894–903.

Smith, S. & Somerton, G. (1981), 'Ms 1981. strap: A user-oriented computer analysis system for groundfish research vessel survey data', Can. Tech. Rep. Fish. Aquat. Sci **1030**.

Stevenson, S. (1983), 'A review of sampling commercial groundfish catches in newfoundland', Sampling Commercial Catches of Marine Fish and Invertebrates. Canadian Special Publication of Fisheries and Aquatic Sciences **66**, 29–38.

Sullivan, P. J. (1992), 'A kalman filter approach to catch-at-length analysis', Biometrics pp. 237–257.

Team, S. D. et al. (2016), 'Stan modeling language users guide and reference manual', Technical report .

Thorson, J. T., Fonner, R., Haltuch, M. A., Ono, K. & Winker, H. (2016), 'Accounting for spatiotemporal variation and fisher targeting when estimating abundance from multispecies fishery data', Canadian Journal of Fisheries and Aquatic Sciences **74**(11), 1794–1807.

Thorson, J. T., Johnson, K. F., Methot, R. D. & Taylor, I. G. (2017), 'Model-based estimates of effective sample size in stock assessment models using the dirichlet-multinomial distribution', Fisheries Research **192**, 84–93.

Thorson, J. T., Shelton, A. O., Ward, E. J. & Skaug, H. J. (2015), 'Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for west coast groundfishes', ICES Journal of Marine Science **72**(5), 1297–1310.

Tomlinson, P. K. (1970), 'A generalization of the murphy catch equation', Journal of the Fisheries Board of Canada **27**(4), 821–825.

Trijoulet, V., Fay, G. & Miller, T. J. (2019), 'Performance of a state-space multispecies model: What are the consequences of ignoring predation and process errors in stock assessments?', Journal of Applied Ecology .

Trijoulet, V., Holmes, S. J. & Cook, R. M. (2017), 'Grey seal predation mortality on three depleted stocks in the west of scotland: What are the implications for stock assessments?', Canadian Journal of Fisheries and Aquatic Sciences **75**(5), 723–732.

Vincent, M., G. P. & Hampton, J. (2019), 'Stock assessment of skipjack tuna in the western and central pacific ocean', Scientific Committee fifteenth Regular Session Bali, Indonesia .

Weijerman, M., Fulton, E. A. & Brainard, R. E. (2016), 'Management strategy evaluation applied to coral reef ecosystems in support of ecosystem-based management', PLoS One **11**(3).

Wheeland, L., Dwyer, K., Morgan, M., Rideout, R. & Rogers, B. (2018), 'Assessment of american plaice in div. 3lno', Évaluation .

Xiao, Y. & Wang, Y.-G. (2007), 'A revisit to pope's cohort analysis', Fisheries research **86**(2-3), 153–158.

Appendix A

Derivation of density function for second phase length-at-age data

Without loss of generality, we assume that $l \in S_k$. It follows from conditional probability that,

$$\begin{aligned} f(l, a|LSAS, \boldsymbol{\theta}) &= \sum_k [f(l, a|l \in S_k; LSAS, \boldsymbol{\theta}) Q_k(LSAS, \boldsymbol{\theta})] \\ &= f(l, a|l \in S_k; LSAS, \boldsymbol{\theta}) Q_k(LSAS, \boldsymbol{\theta}), \end{aligned}$$

where $Q_k(LSAS, \boldsymbol{\theta})$ is the probability for a second phase individual to be in length stratum S_k. Since the subsampling in each stratum is random, $f(l, a|l \in S_k; LSAS, \boldsymbol{\theta}) = f(l, a|l \in S_k; \boldsymbol{\theta})$, so we can write,

$$\begin{aligned} f(l, a|LSAS, \boldsymbol{\theta}) &= f(l, a|l \in S_k; \boldsymbol{\theta}) Q_k(LSAS, \boldsymbol{\theta}) \\ &= \frac{f(l, a|\boldsymbol{\theta})}{Q_k(\boldsymbol{\theta})} Q_k(LSAS, \boldsymbol{\theta}). \end{aligned}$$

Appendix B

Horvitz-Thompson weights

For the calibrated weighted (CW) method, weights are defined as $w_i = g_i \frac{N_{k(l)}}{n_{k(l)}}$ for $i = 1, ..., N$, where $k(l)$ represents length l in stratum k, such that the difference between w_i and $\frac{N_{k(l)}}{n_{k(l)}}$ is minimized subject to some constraints. To calibrate the first order moment (the mean), the weights w_i are found by minimizing the chi-square distance $G(w_i, d_i) = \frac{(w_i - d_i)^2}{2d_i}$, subject to

$$\sum_{i=1}^{N} \gamma_i w_i l_i = \sum_{i=1}^{N} l_i.$$

Here, γ_i is a binary indicator of whether or not the ith subject is fully observed. These new calibrated weights can be found using the method of Lagrange multipliers, and can easily be implemented using equation (9) from Breslow et al. (2009). From that equation it follows that,

$$g_i = 1 - \lambda_c l_i,$$

where

$$\lambda_c = \frac{\sum_{i=1}^{n} \frac{N_{k(l)}}{n_{k(l)}} l_i - \sum_{i=1}^{N} l_i}{\sum_{i=1}^{n} l_i^2},$$

then

$$w_i = (1 - \lambda_c l_i) \frac{N_{k(l)}}{n_{k(l)}}.$$

Appendix C

Neglecting spatial structure of sampling scheme

Assume that from a model M, we generated a random sample A with the individuals enumerated as $a_1, a_2, \ldots, a_N$. In an arbitrary length stratum k, there are N_k individuals of sample A which are denoted as $a_{k,1}, a_{k,2}, \ldots, a_{k,N_k}$. Now we make a random permutation of sample A, which gives sample B with individuals $b_1, b_2, \ldots, b_N$, and the N_k individuals in stratum k are now denoted as $b_{k,1}, b_{k,2}, \ldots, b_{k,N_k}$. The individuals in sample B are allocated into N_s survey stations, and length stratified random sampling is conducted in each stratum with different rules. The final result is to obtain n_k individuals in stratum k for age measurement. Due to the random permutation between A and B, despite the complicated sampling design, this is in fact a random sample of size n_k from $a_{k,1}, a_{k,2}, \ldots, a_{k,N_k}$ for any stratum k, then all the inference approaches and conclusions of this chapter can be applied. Therefore, we can treat the final length at age data from sample B (or equivalently from sample A, since A and B have identical set of individuals) as from standard LSAS defined by (2). Equivalent to A, sample B is also a random sample generated from model M. We reach the conclusion that for any random sample generated from model M, we can neglect the complicated station-wise sampling design, and treat the final data as from standard LSAS.

Appendix D

Supporting figures for state-space stock assessment model for 3LNO American plaice

Figure D.1: Model predicted survey fits (red) and observed survey indices at age (black) for ages 1-7 for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO.

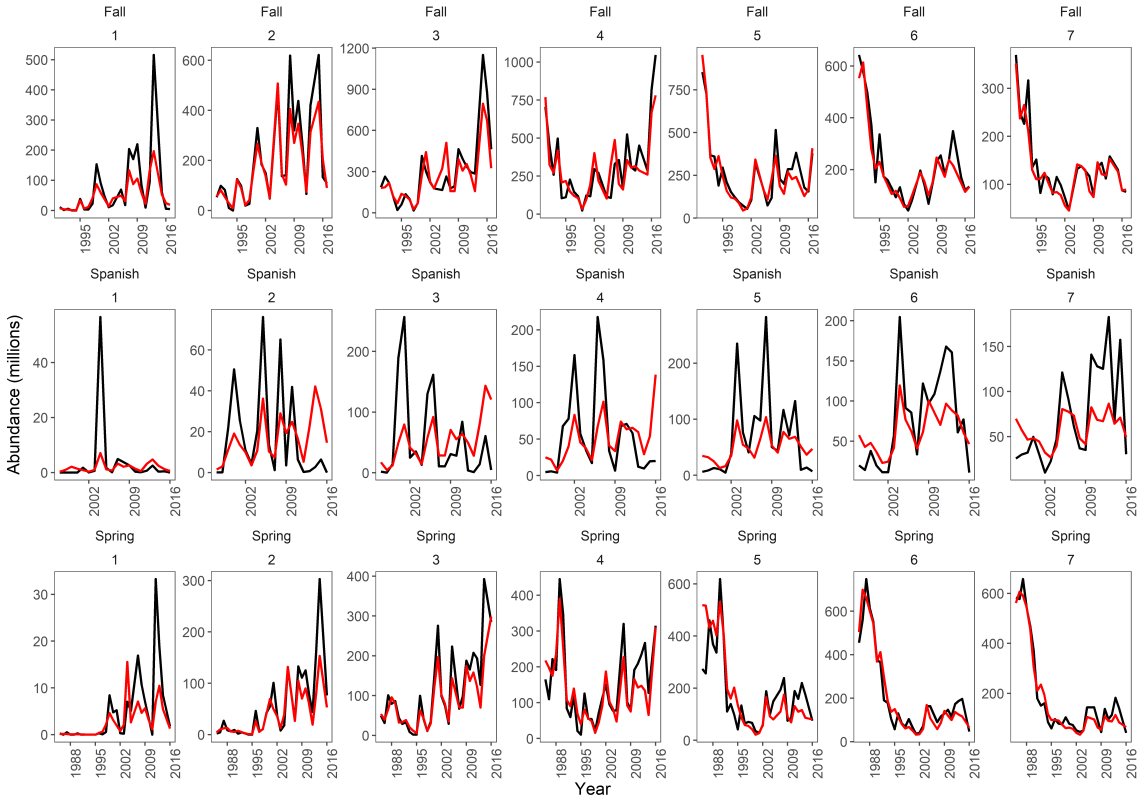


Figure D.2: Model predicted survey fits (red) and observed survey indices at age (black) for ages 8+ for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO.

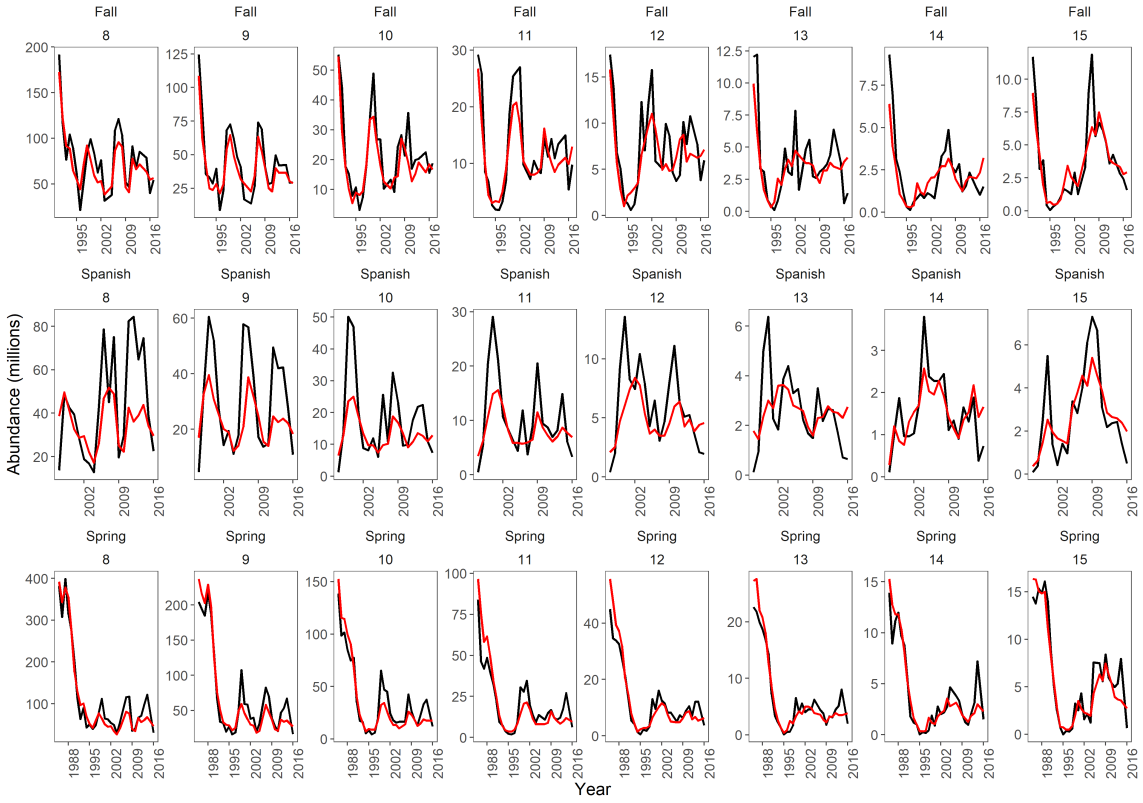


Figure D.3: Standardized bubble plot of residuals at age for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative.

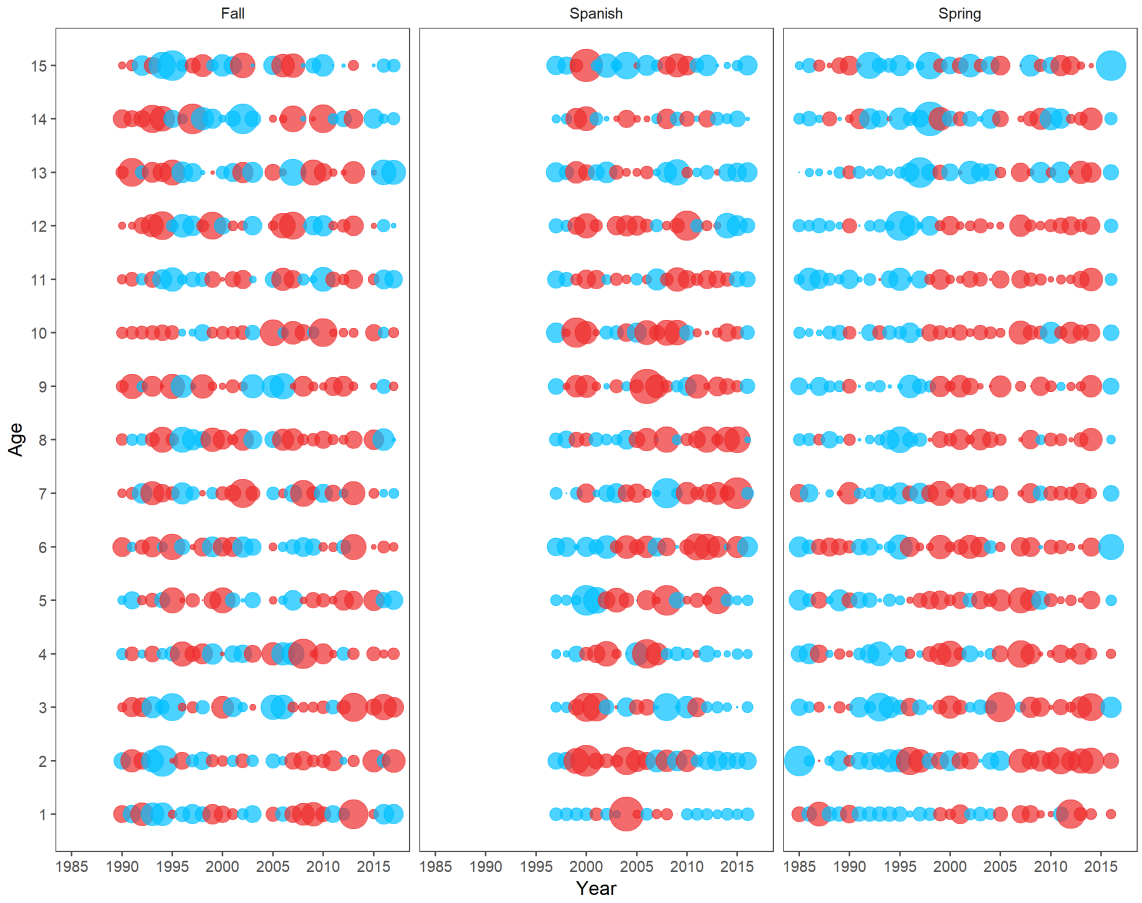


Figure D.4: Standardized residuals for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO.

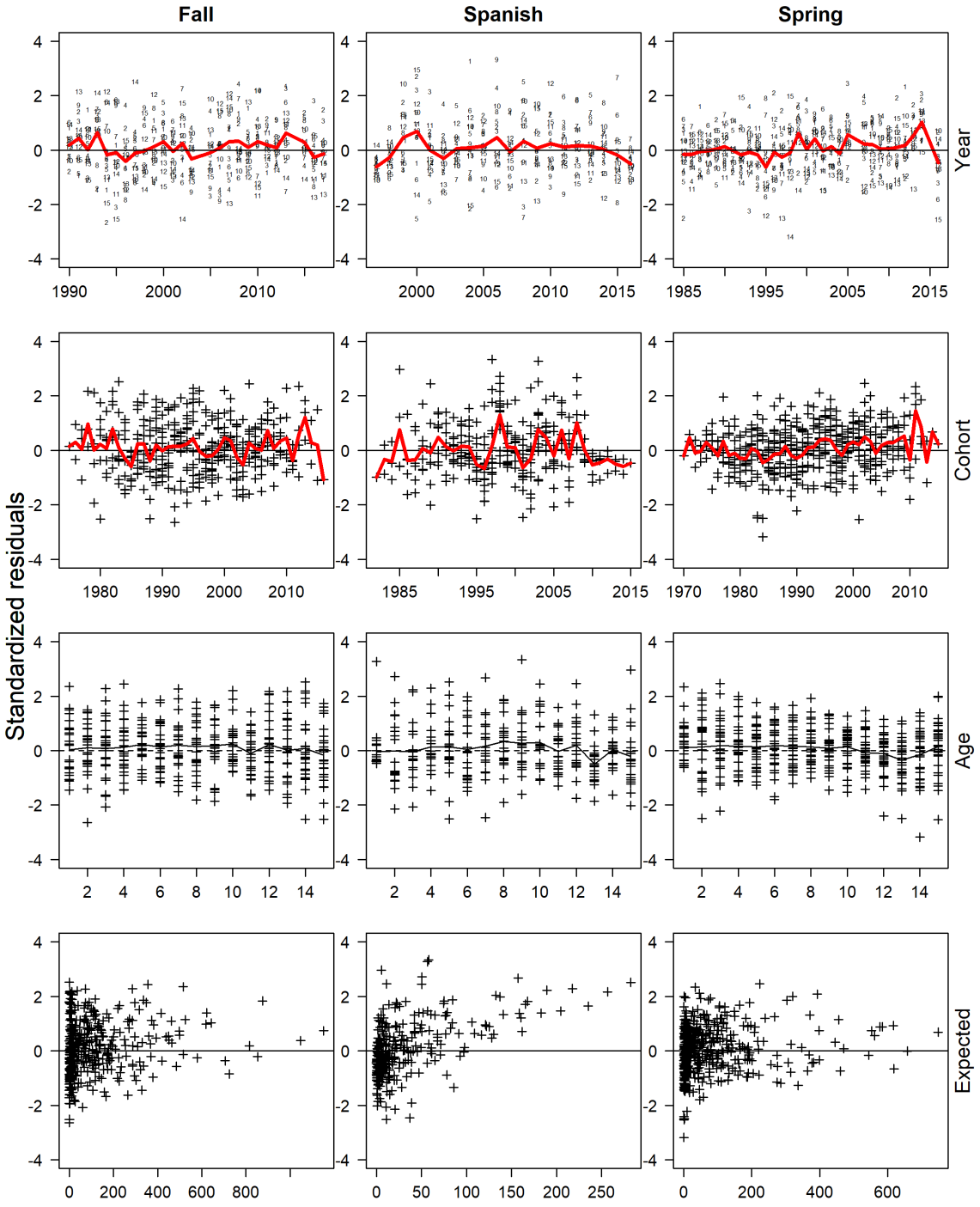


Figure D.5: Model predicted catch proportions at age (red) vs observed catch proportion at age (black) for American plaice in NAFO Divisions 3LNO.

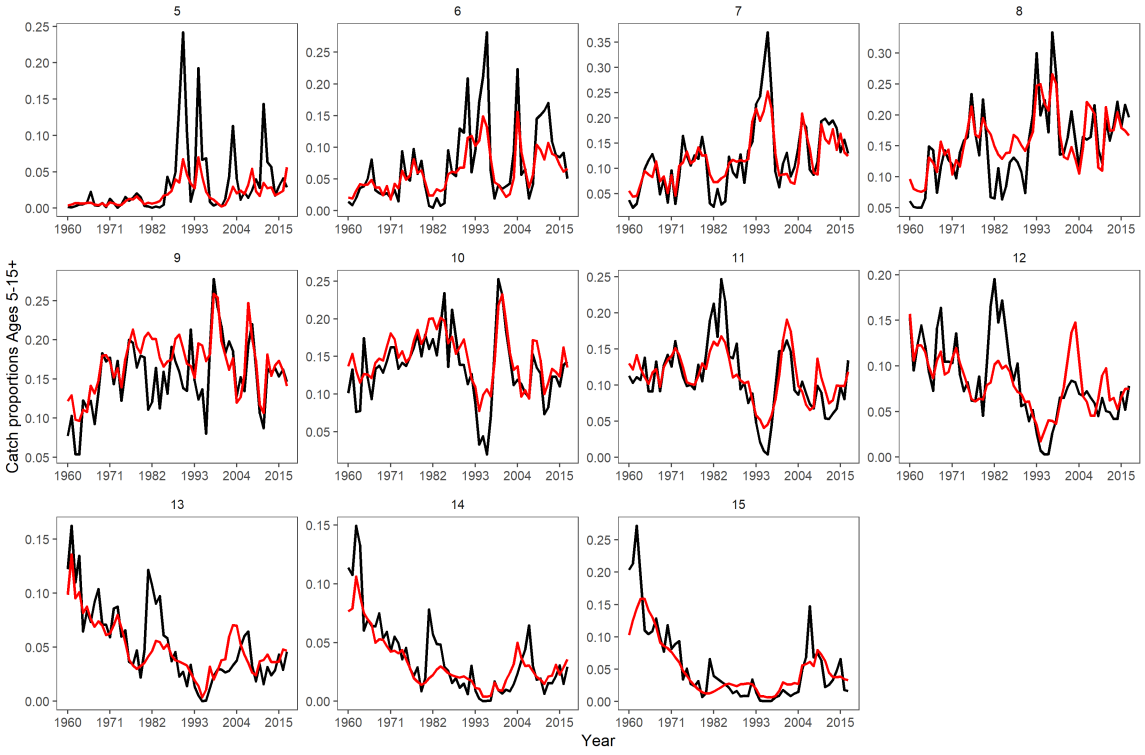


Figure D.6: Standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO.

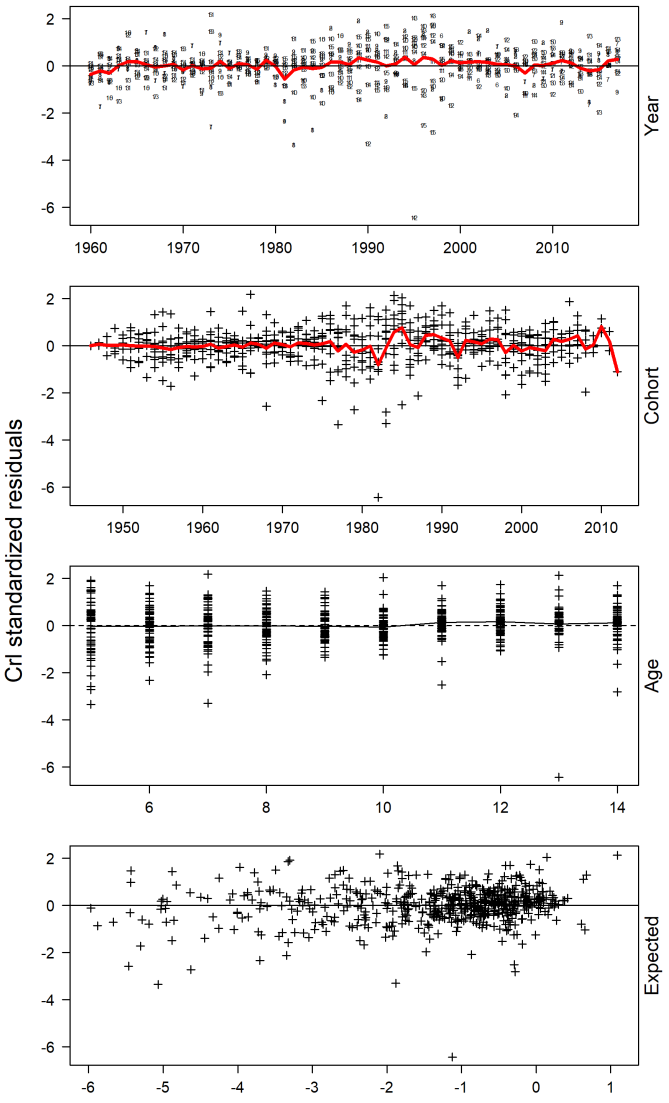


Figure D.7: Bubble plot of standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO.

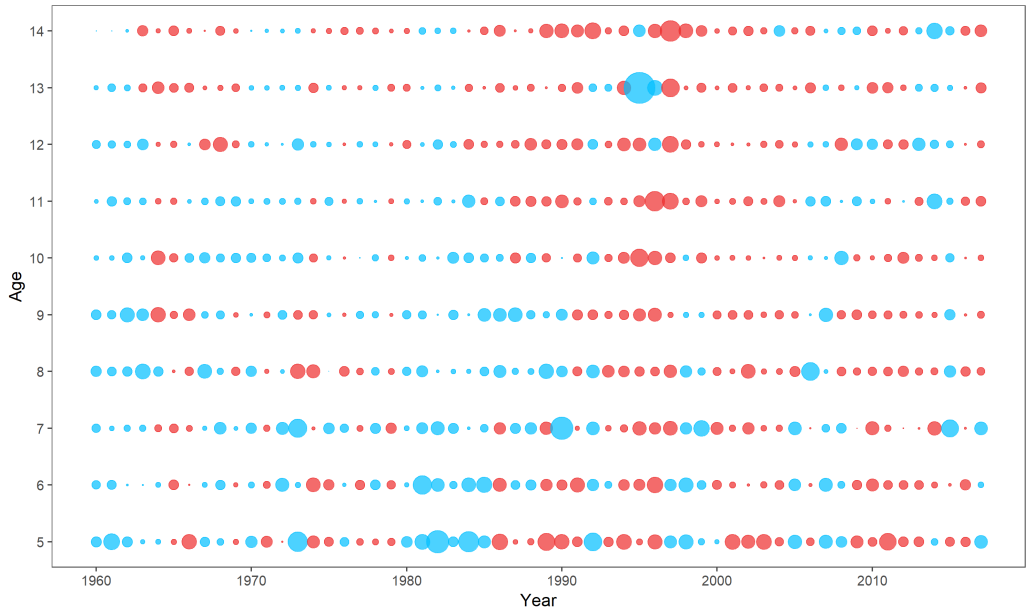


Figure D.8: Model predicted log fishing mortality rates for American plaice in NAFO Divisions 3LNO.

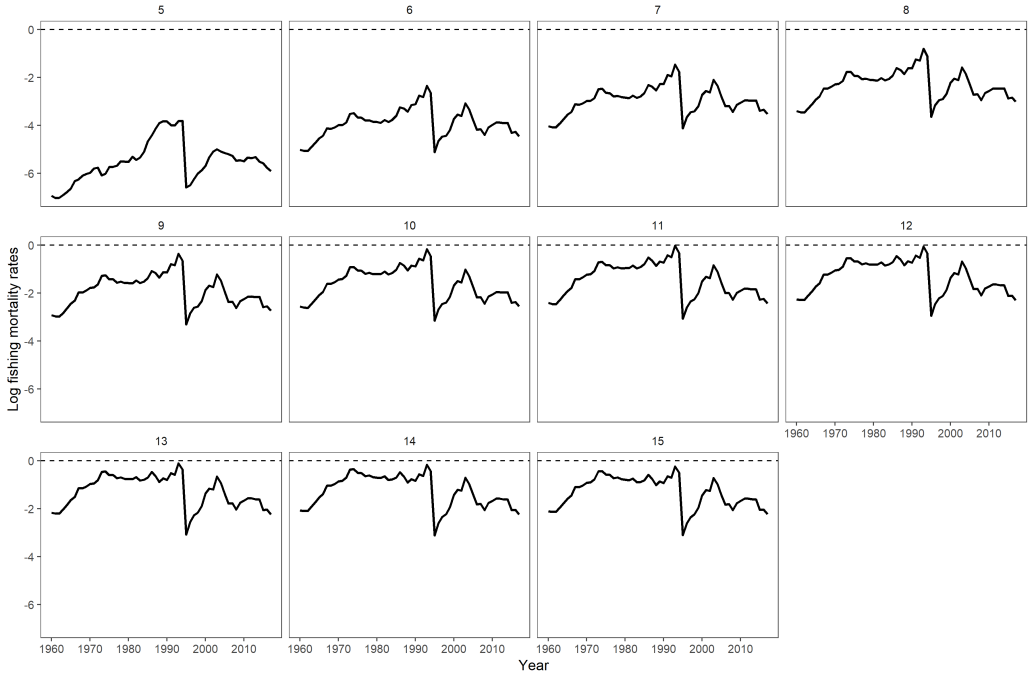


Figure D.9: Model predicted log fishing mortality deviations for American plaice in NAFO Divisions 3LNO.

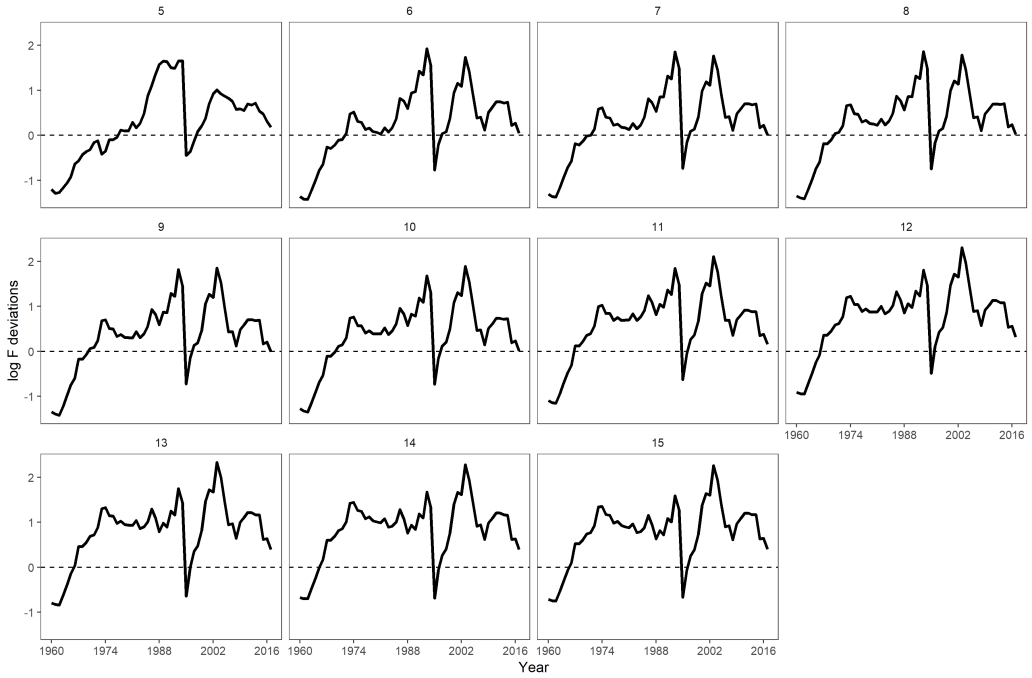
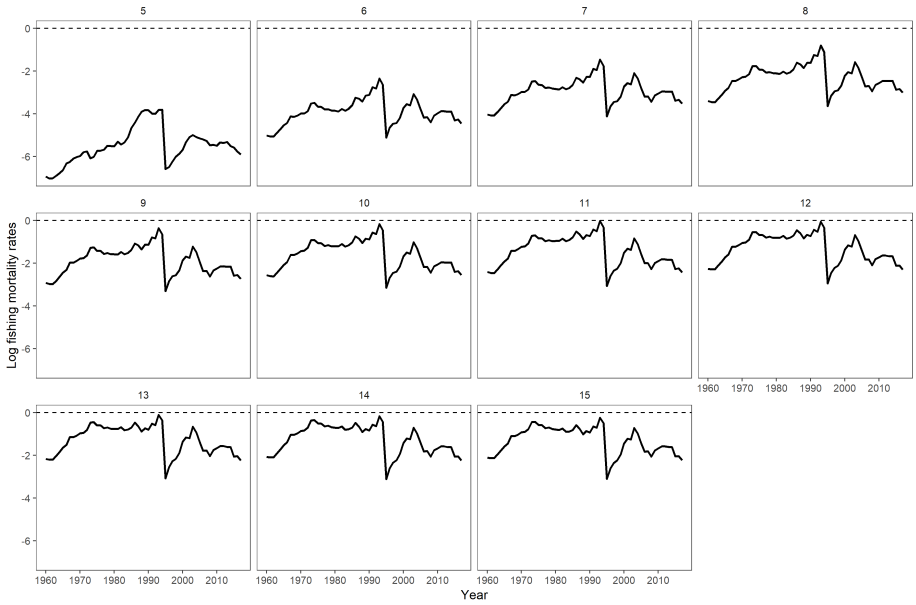


Figure D.10: Model predicted fishing mortality rates for American plaice in NAFO Divisions 3LNO.



Appendix E

Data inputs for state-space stock assessment model for 3LNO American plaice

Figure E.1: Abundance data for ages 1-7 for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO.

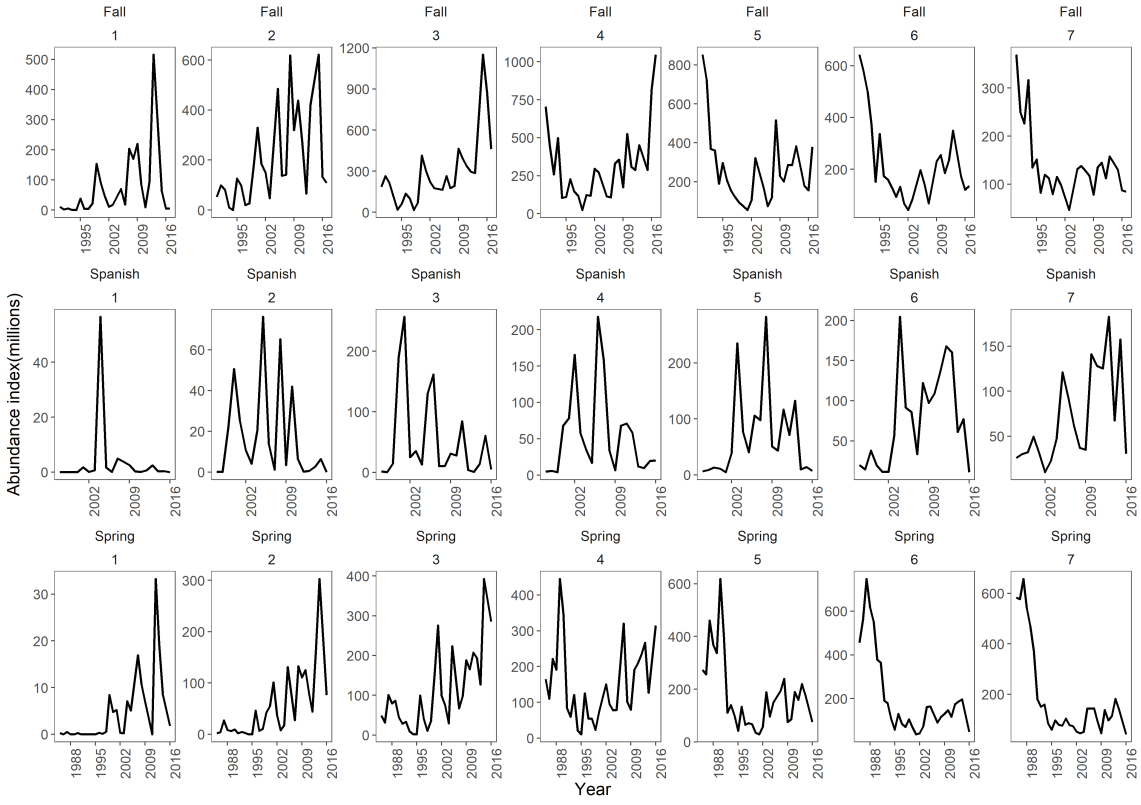


Figure E.2: Abundance data for ages 8+ for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO.

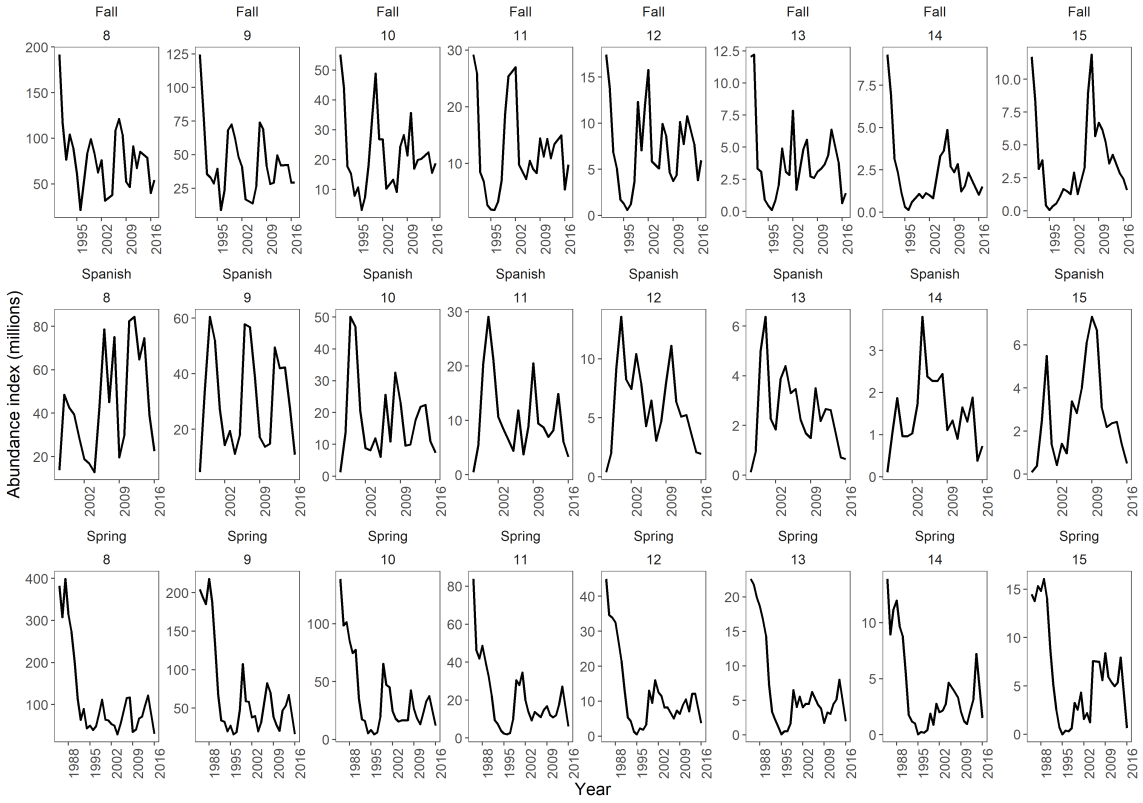


Figure E.3: Standardized proportion by age across years for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative and black is zeros

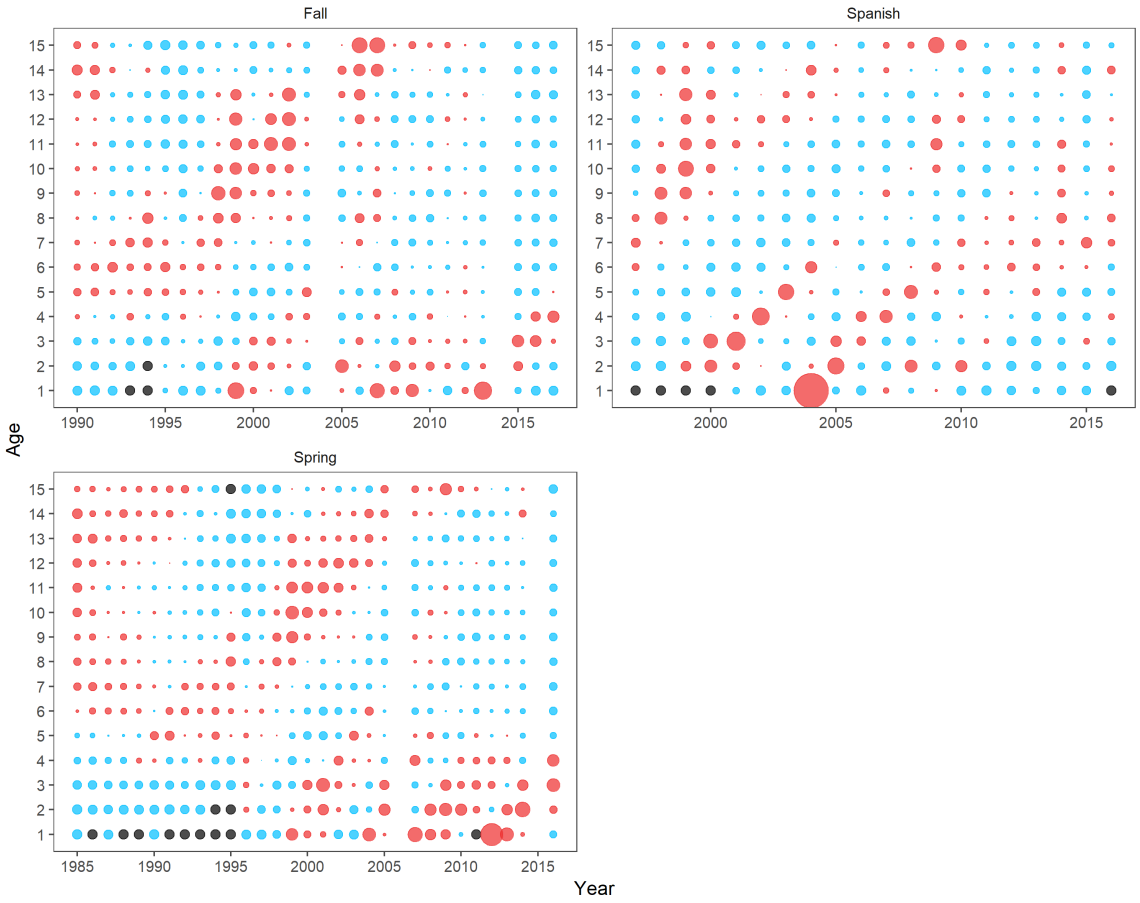


Figure E.4: Mean standardized survey indices for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. Red is fall, green is Spanish and blue is spring survey.

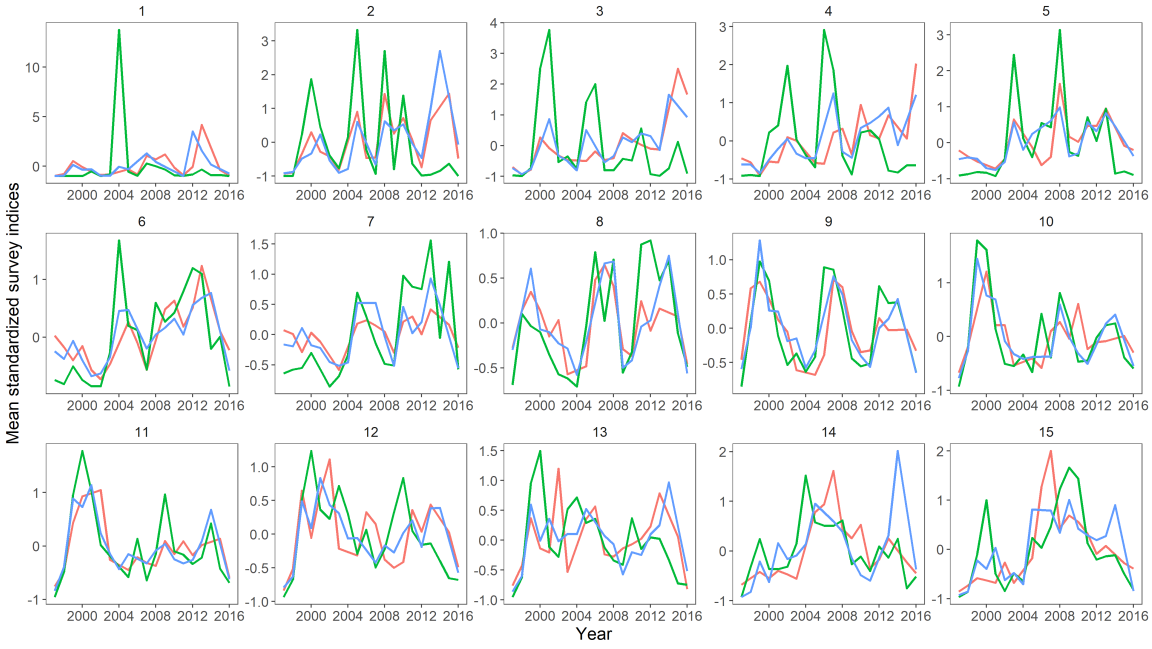


Figure E.5: Landings estimates for American plaice in NAFO Divisions 3LNO.

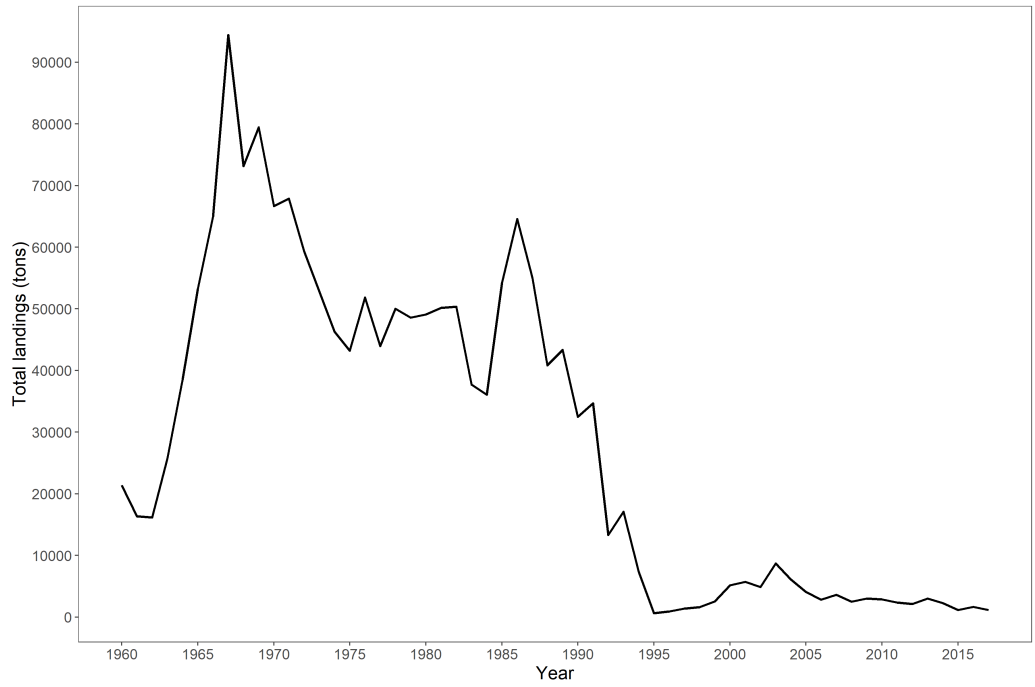


Figure E.6: Catch proportions at age for American plaice in NAFO Divisions 3LNO.

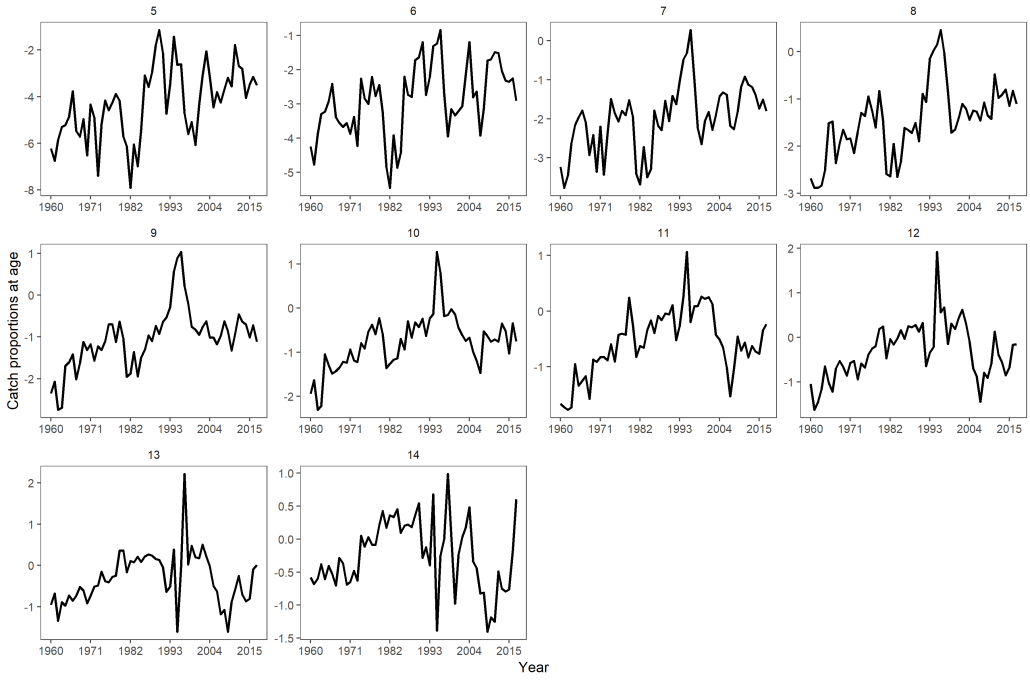


Figure E.7: Proportion mature at age for American plaice in NAFO Divisions 3LNO.

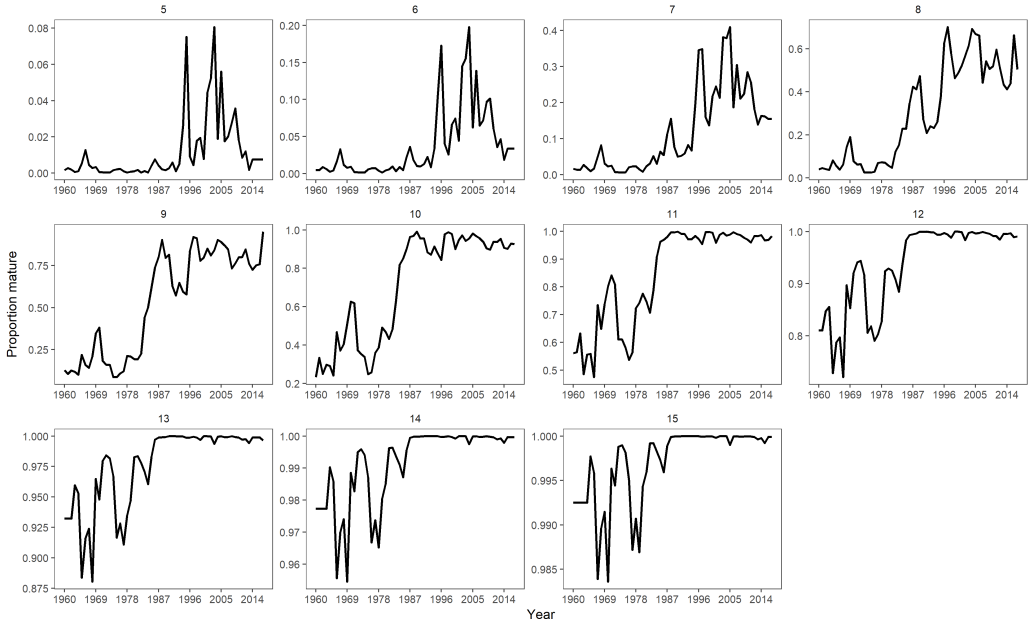


Figure E.8: Stock weights at age for American plaice in NAFO Divisions 3LNO.

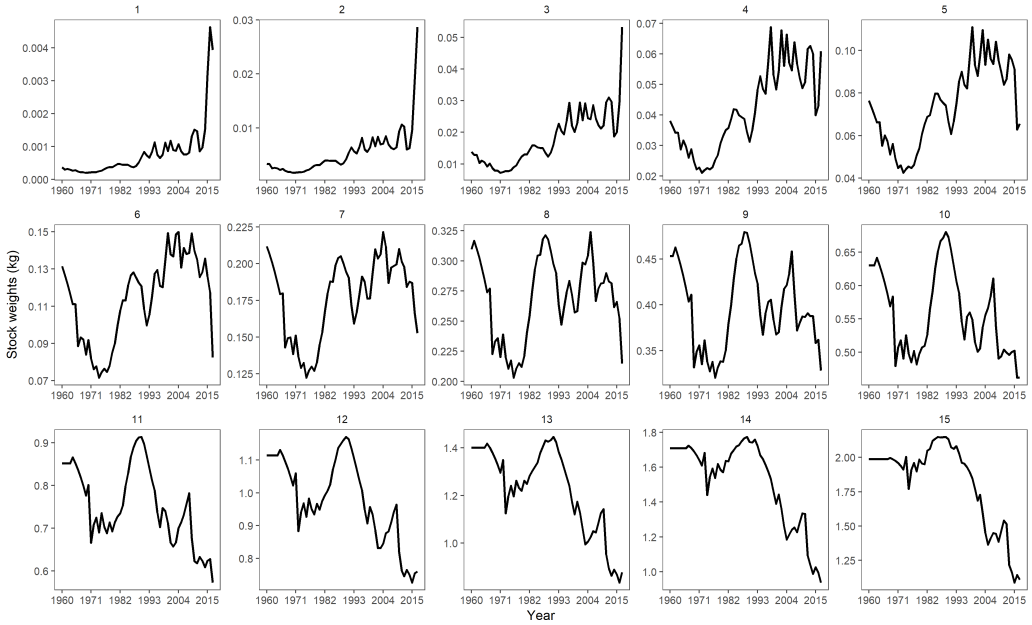
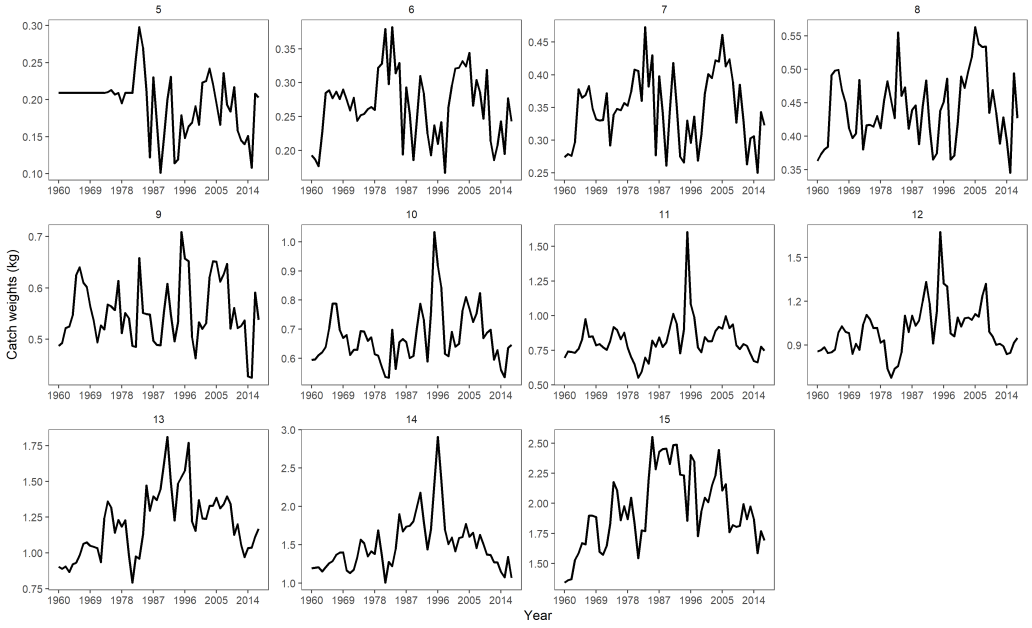


Figure E.9: Catch weights at age for American plaice in NAFO Divisions 3LNO.



Appendix F

Exploratory model (EM) fitting

EM1

The process model is identical to the process model described in 3.2.1. The observation model differs from that described in 3.2.2 by: 1) one crl variance parameter estimated for all ages, 2) one coefficient of variation parameter per survey and 3) no correlations across survey and crls observations; $N(0, cv_s I_{s,a,y})$, $LN(0, \sigma_C)$.

	EM1
nll	4958
nparms	64
AIC	10044
BIC	10366

Figure F.1: EM1 estimated population abundance, spawning stock biomass, average fishing mortality rates (ages 9-14) and recruitment for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent 95% confidence intervals.

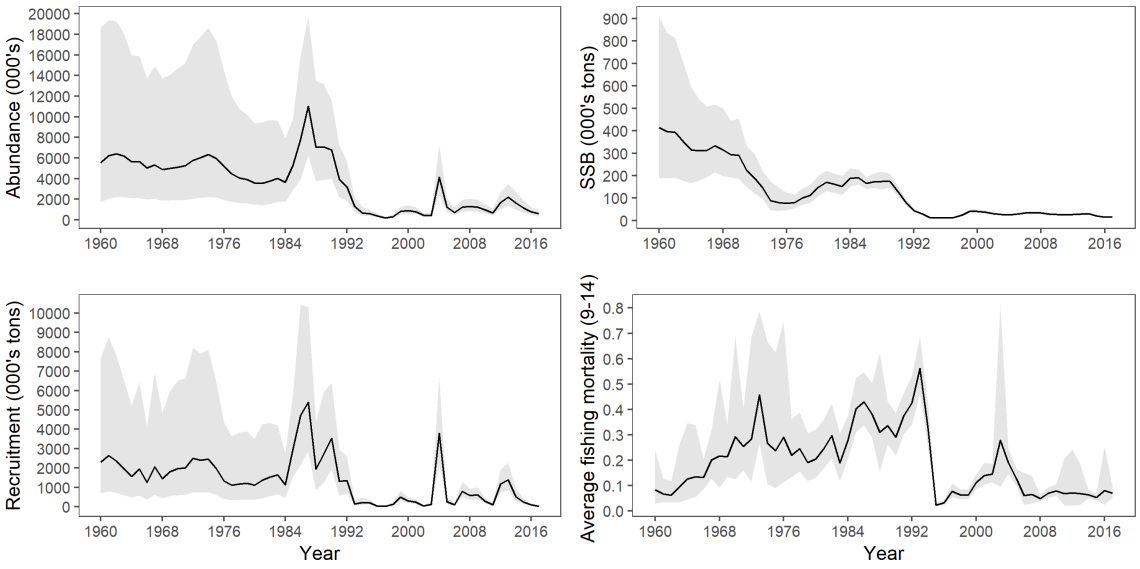


Figure F.2: EM1 predicted survey fits (red) and observed survey indices at age (black) for ages 1-7 for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO.

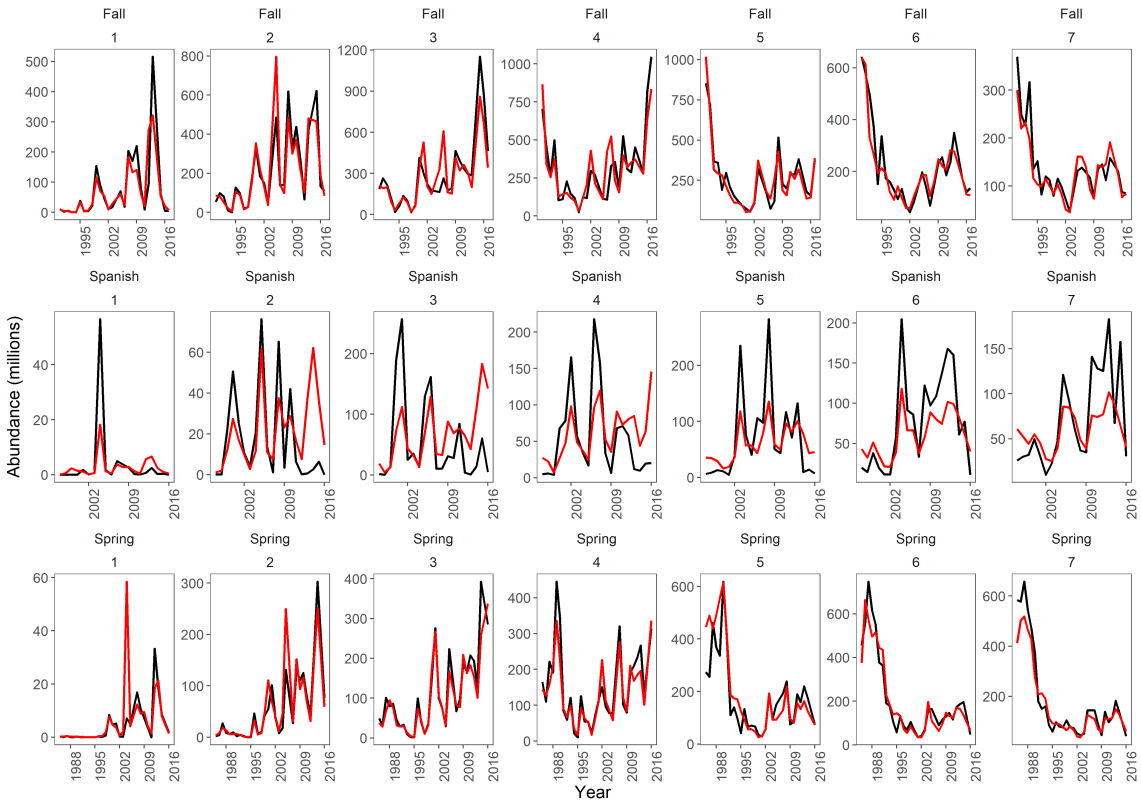


Figure F.3: EM1 predicted survey fits (red) and observed survey indices at age (black) for ages 8+ for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO.

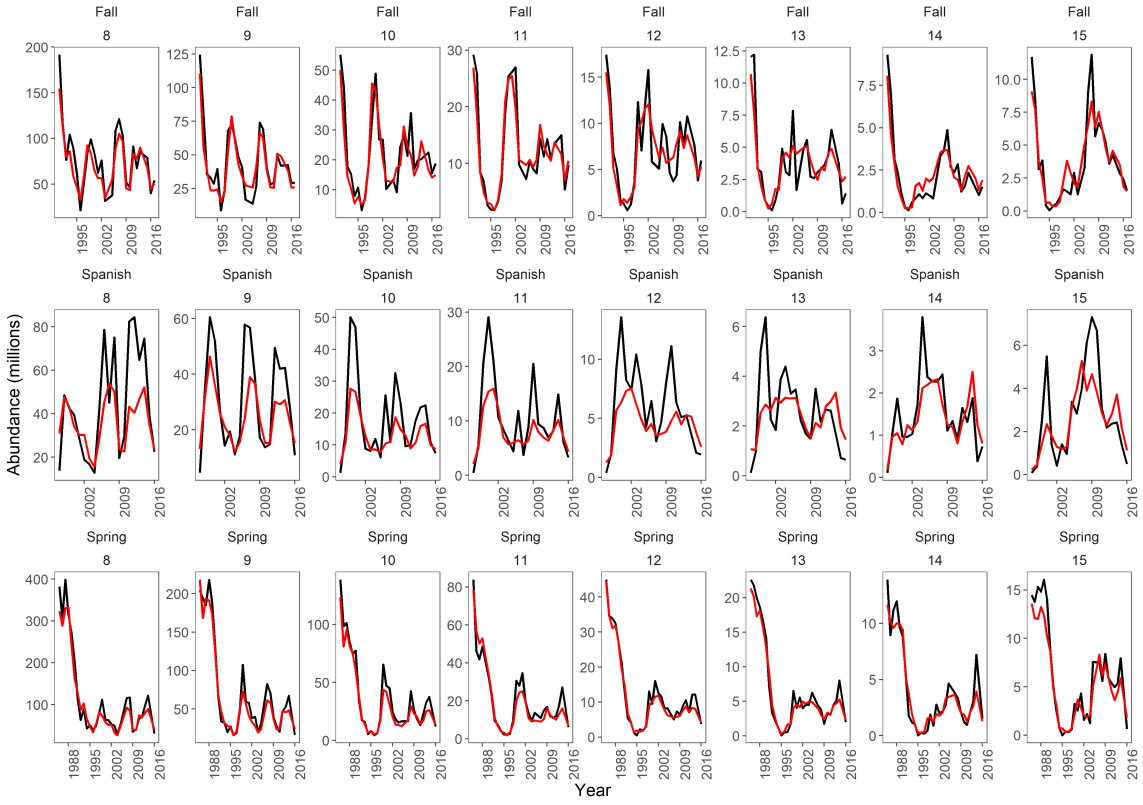


Figure F.4: EM1 standardized bubble plot of residuals at age for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative.

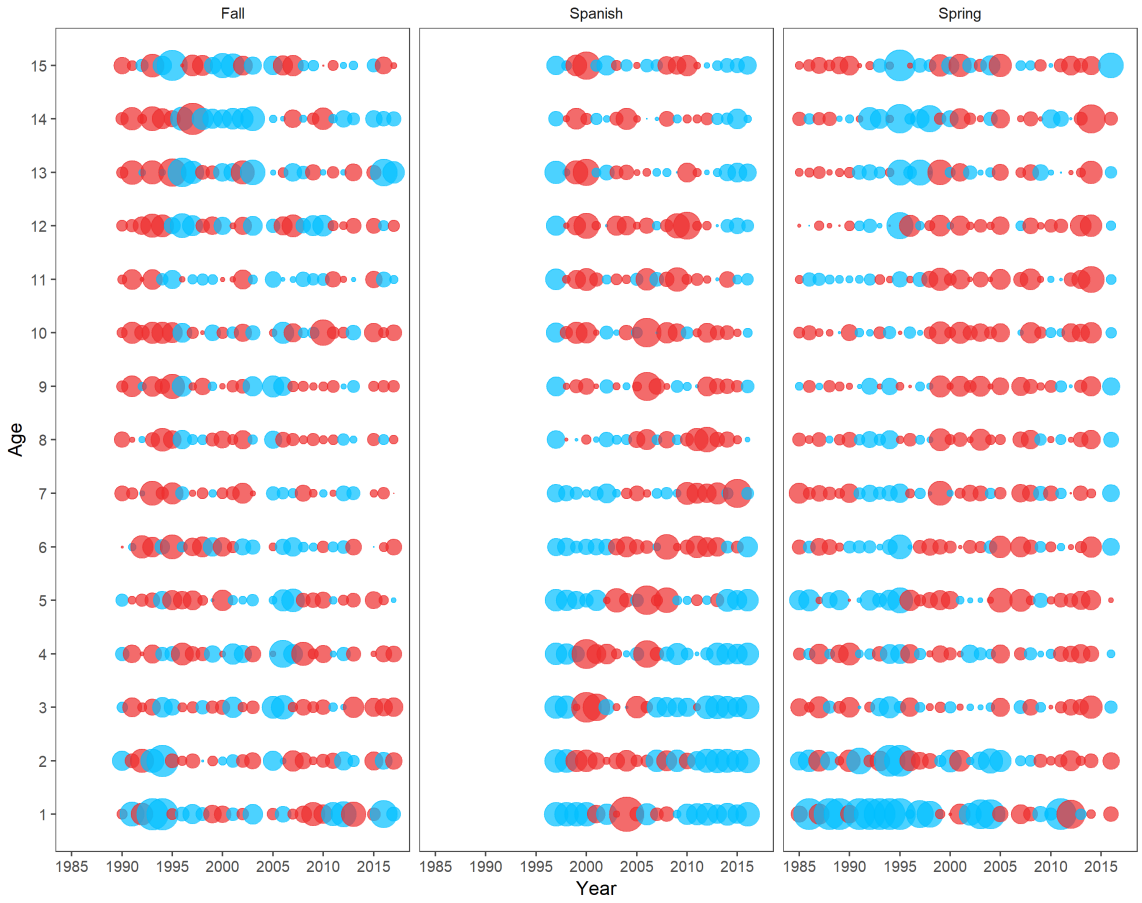


Figure F.5: EM1 standardized residuals for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO.

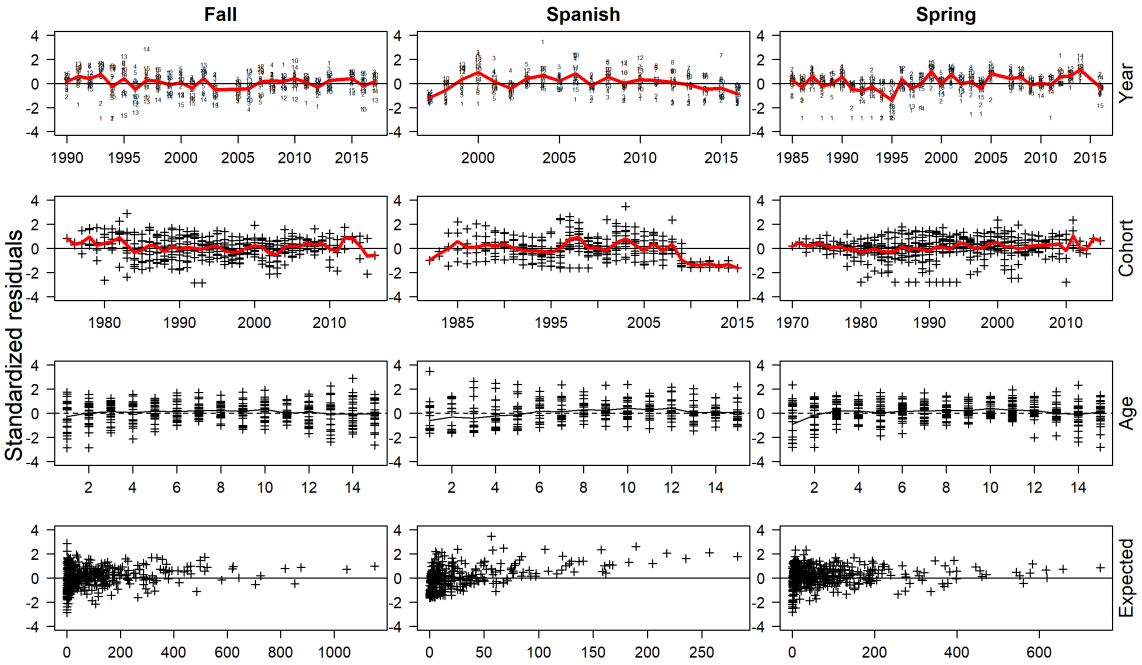


Figure F.6: EM1 estimated log catch numbers for ages 5-15+ (solid line) for American plaice in NAFO Divisions 3LNO, the shaded grey represents the region between the log lower catch bounds and the log upper catch bounds.

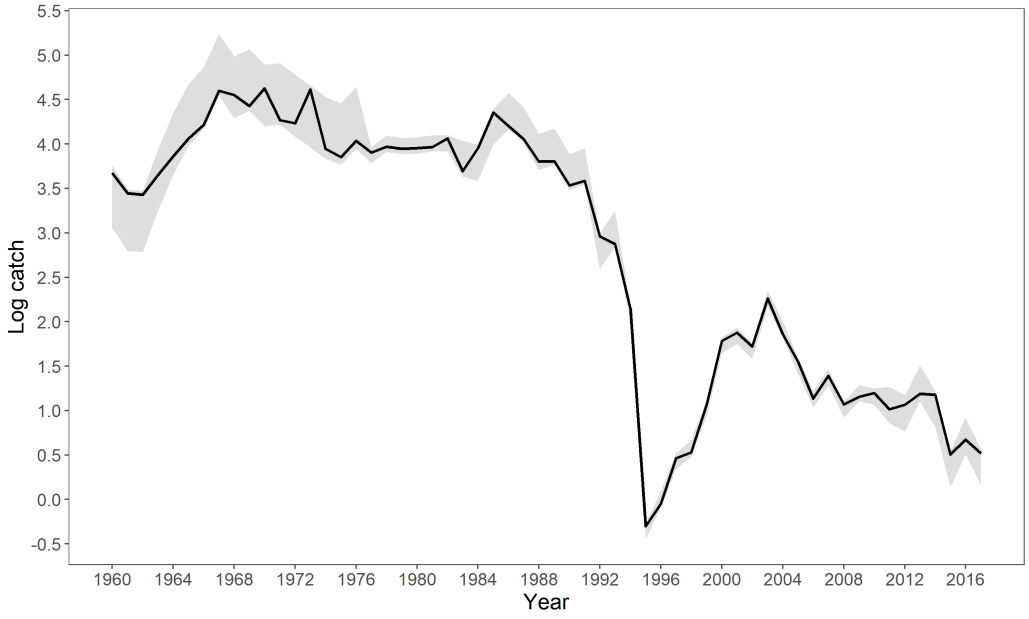


Figure F.7: EM1 predicted catch proportions at age (red) vs observed catch proportion at age (black) for American plaice in NAFO Divisions 3LNO.

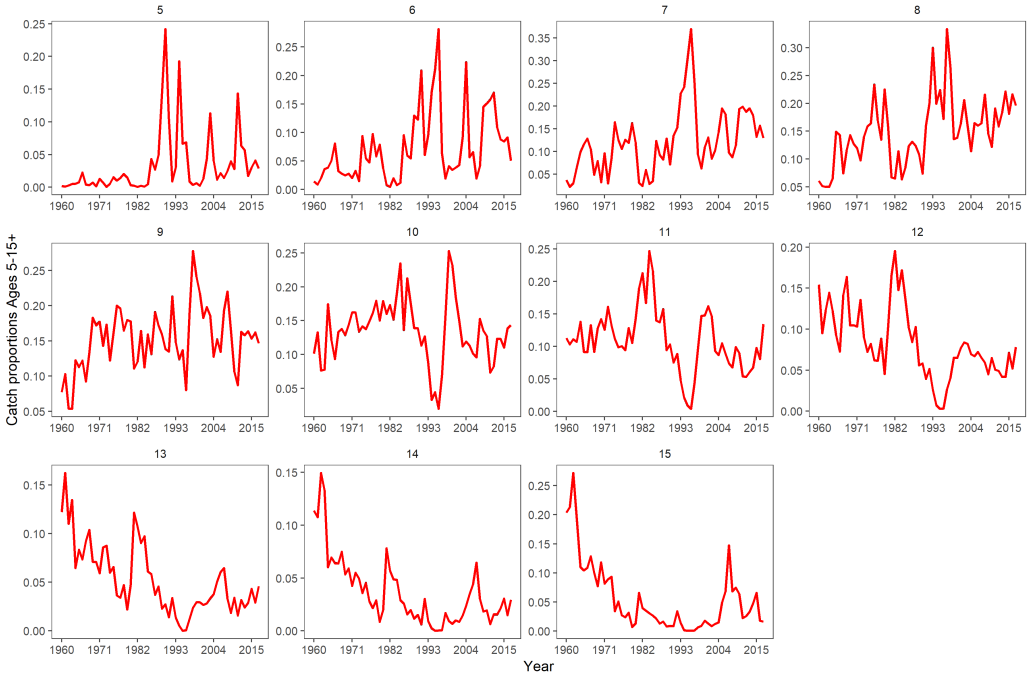


Figure F.8: EM1 standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO.

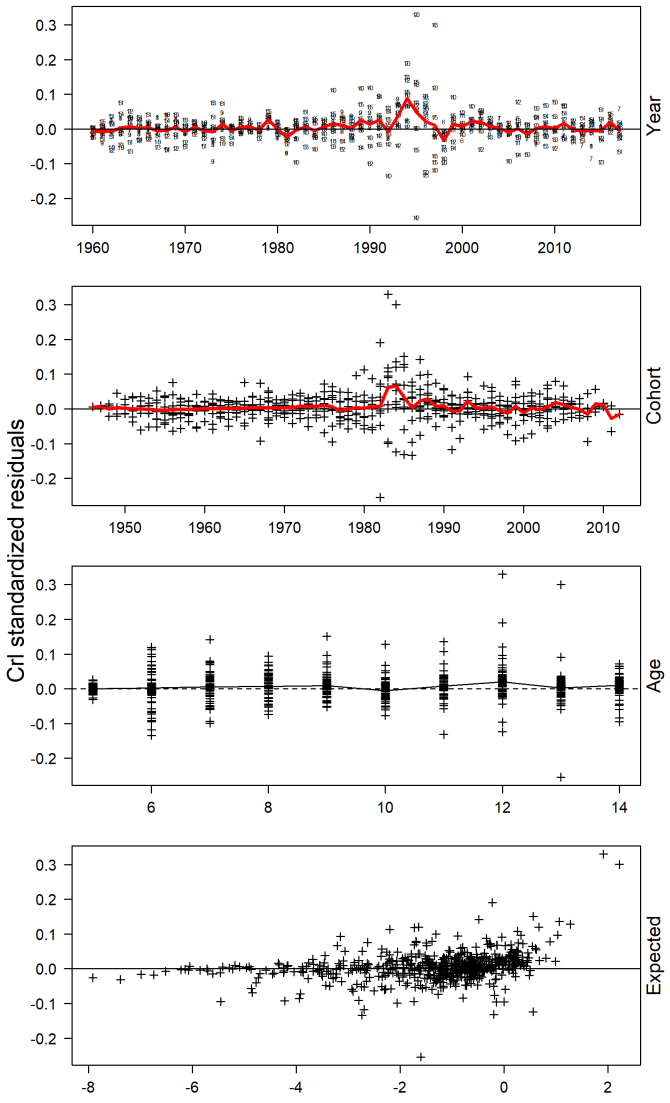


Figure F.9: EM1 bubble plot of standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO.

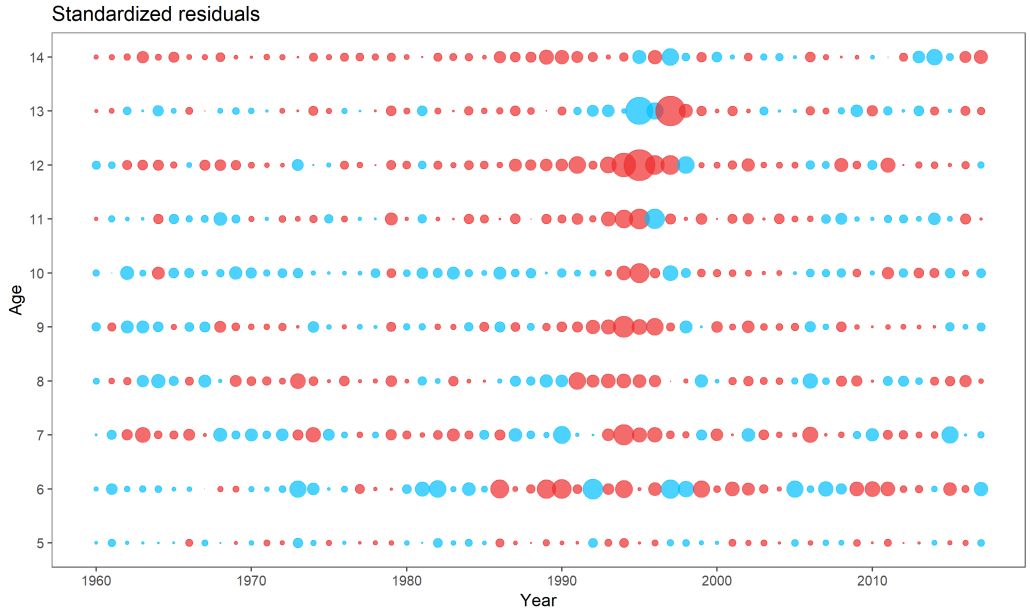


Figure F.10: EM1 predicted log fishing mortality rates for American plaice in NAFO Divisions 3LNO.

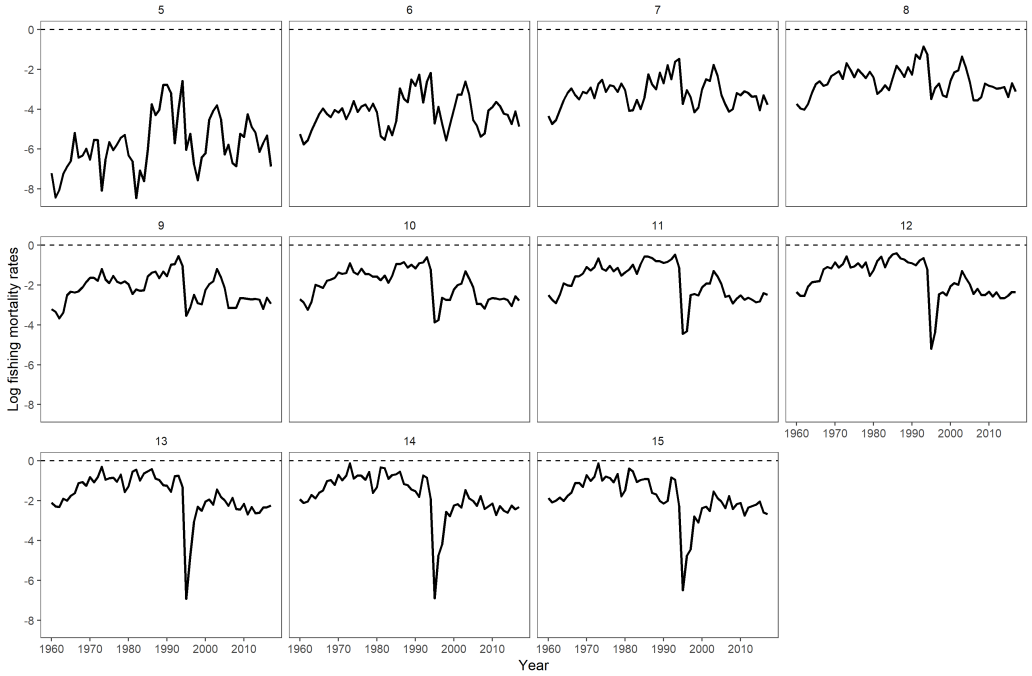


Figure F.11: EM1 predicted log fishing mortality deviations for American plaice in NAFO Divisions 3LNO.

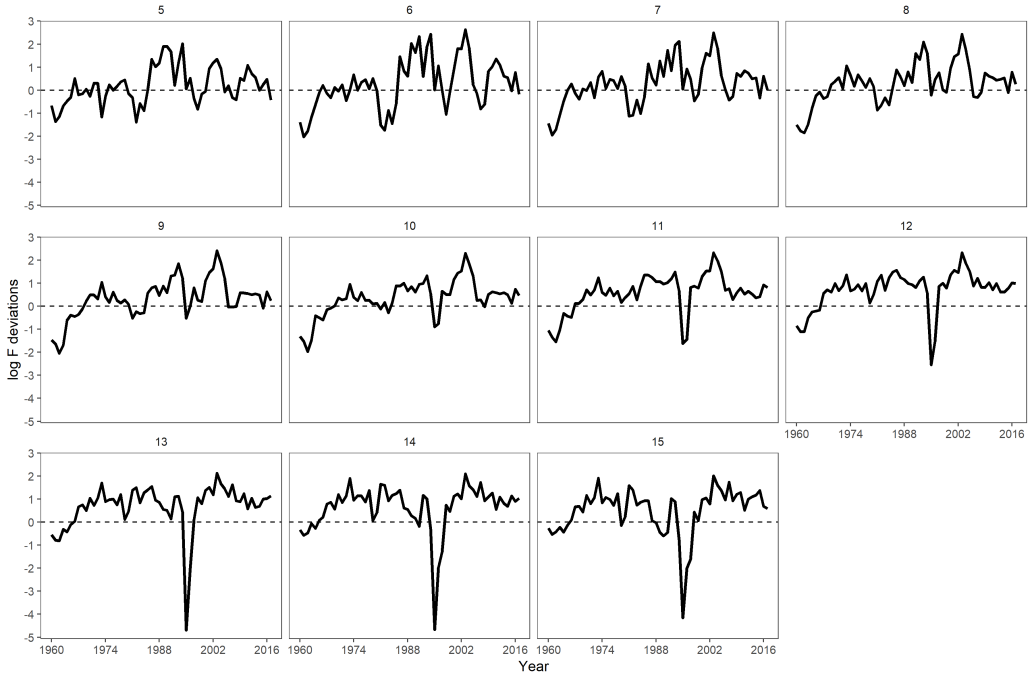


Figure F.12: EM1 predicted fishing mortality rates for American plaice in NAFO Divisions 3LNO.

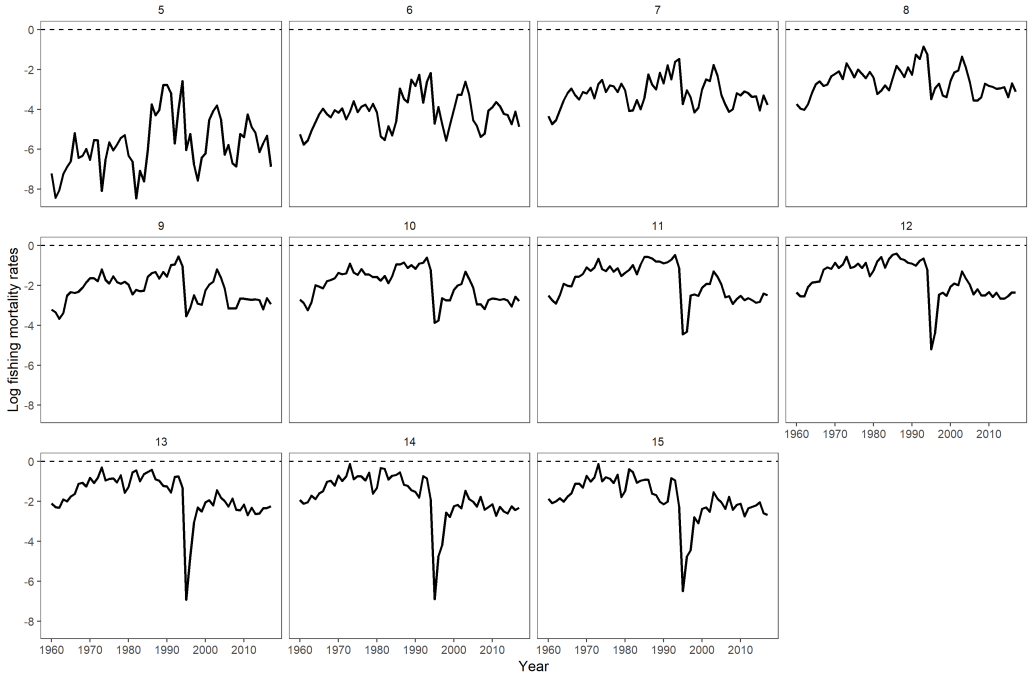


Figure F.13: EM1 predicted process errors at age for American plaice in NAFO Divisions 3LNO.

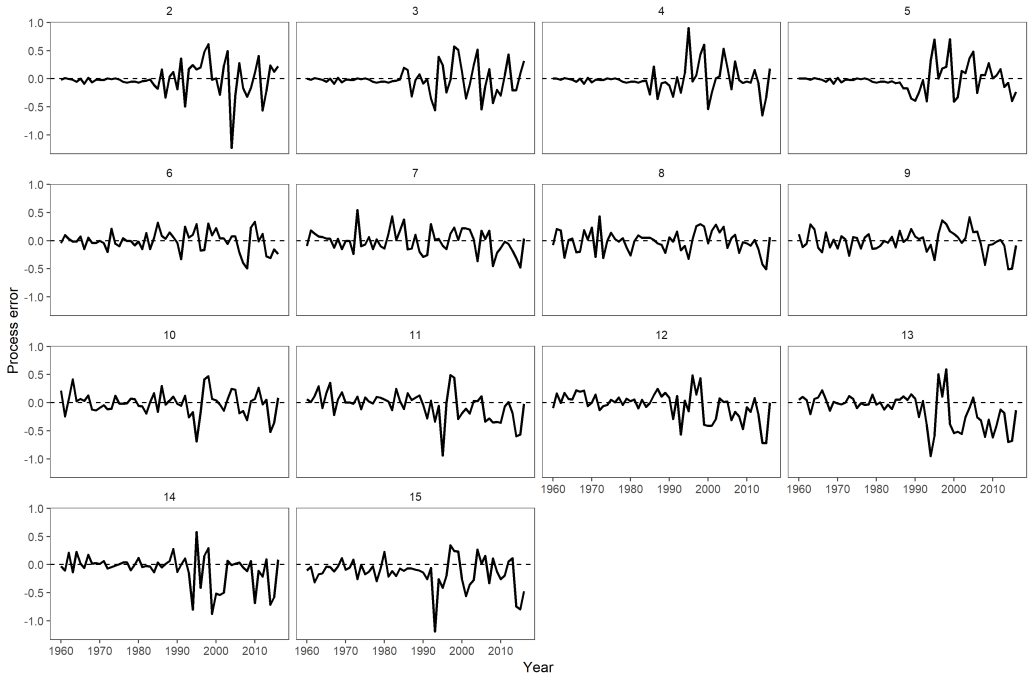


Figure F.14: EM1 bubble plot of predicted process errors at age for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative.

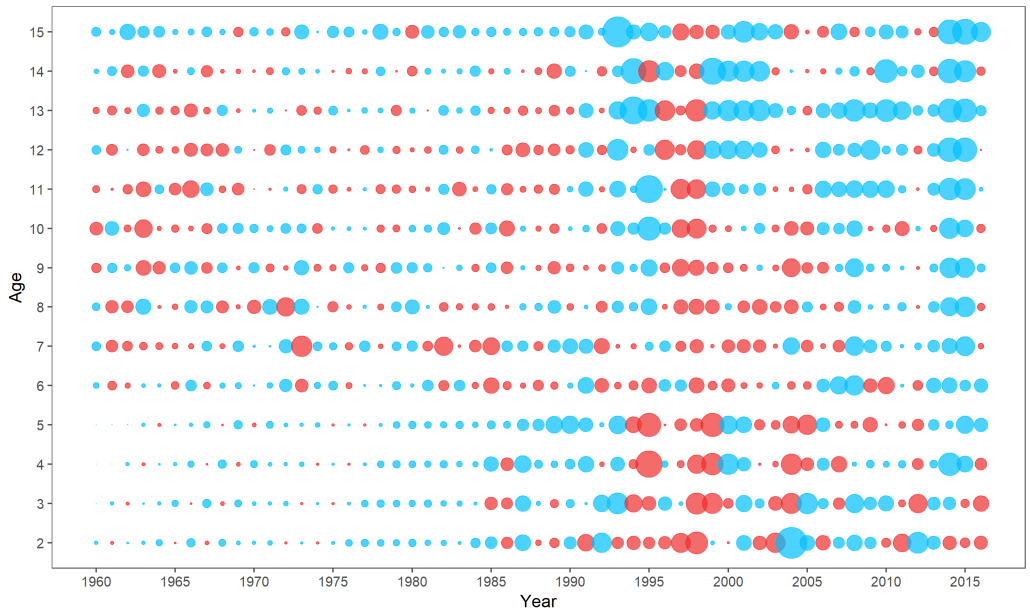


Figure F.15: EM1 survey catchability patterns for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. A separate catchability parameter is estimated for two gear types (Engel and Campelen) for the spring and fall surveys for ages 1-4.

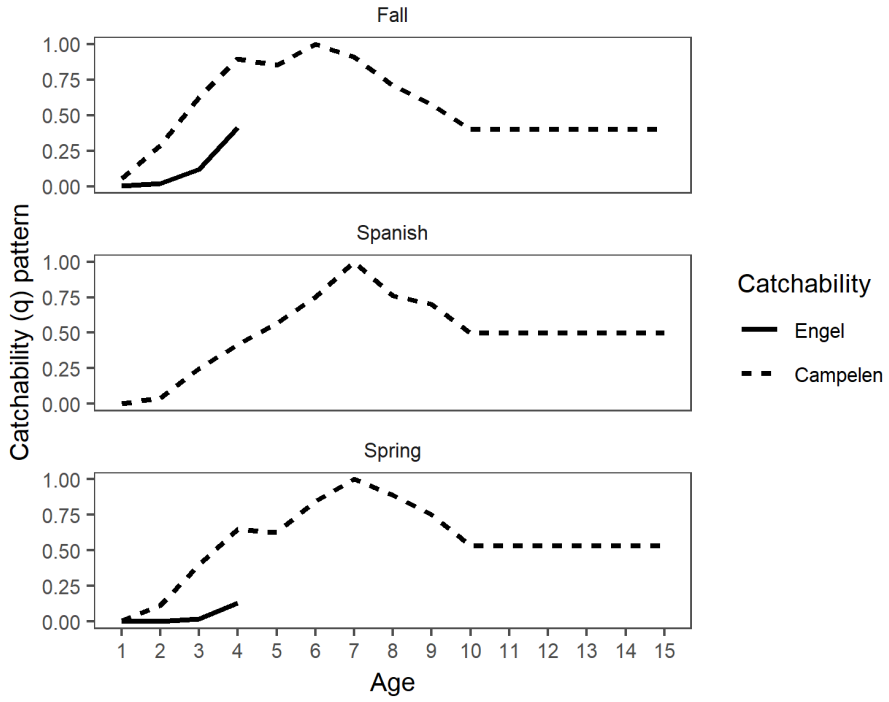


Figure F.16: EM1 (SSM) and VPA estimated spawning stock biomass for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent 95% confidence intervals.

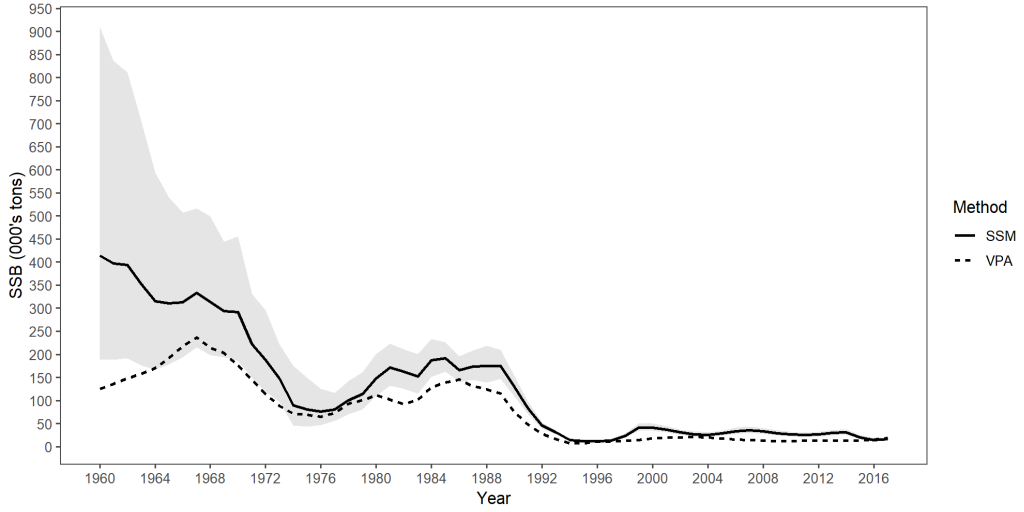
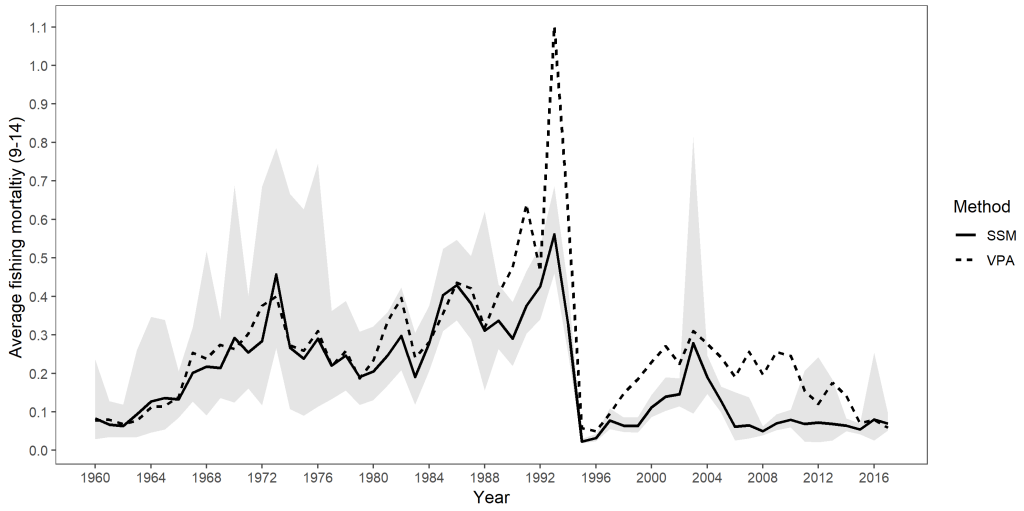


Figure F.17: EM1 (SSM) and VPA estimated average fishing mortality rates (ages 9-14) for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent 95% confidence intervals.



EM2

The process model is identical to the process model described in 3.2.1. The observation model differs from that described in 3.2.2 by: 1) one crl variance parameter estimated for each age, 2) one variance parameter per age per survey and 3) no correlations across survey and crls observations; $N(0, cv_s I_{s,a,y})$, $LN(0, \sigma_C)$.

	EM2
nll	4811
nparms	115
AIC	9853
BIC	10432

Figure F.18: EM2 estimated population abundance, spawning stock biomass (SSB), average fishing mortality rate (ages 9-14) and recruitment for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent 95% confidence intervals.

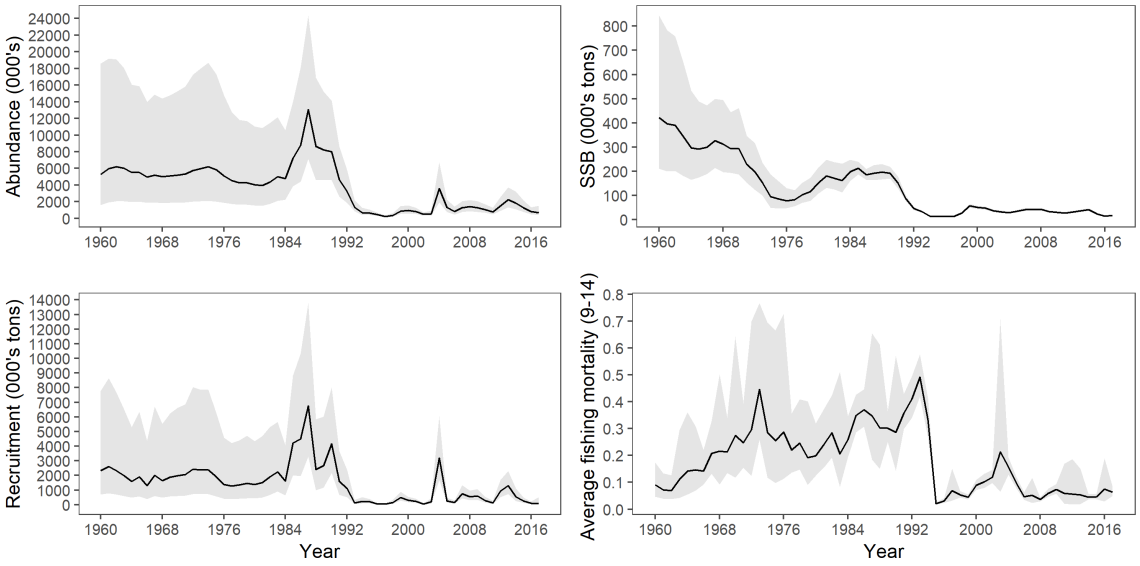


Figure F.19: EM2 predicted survey fits (red) and observed survey indices at age (black) for ages 1-7 for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO.

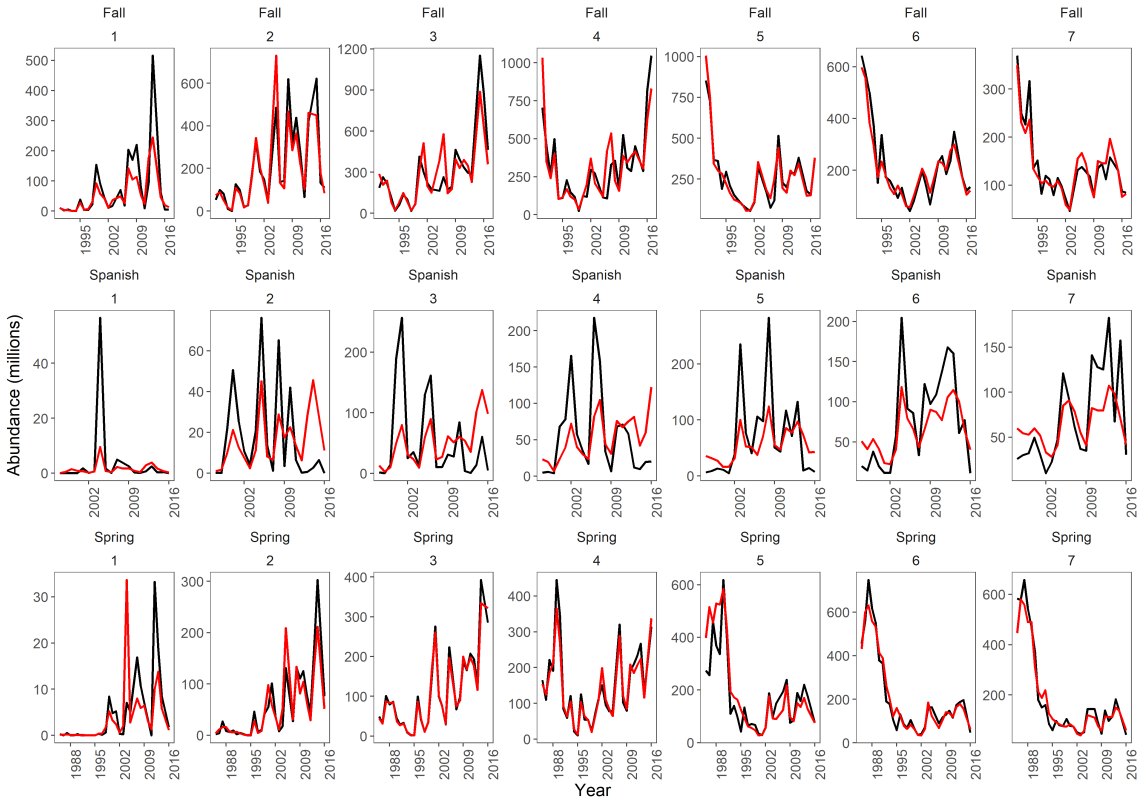


Figure F.20: EM2 predicted survey fits (red) and observed survey indices at age (black) for ages 8+ for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO.

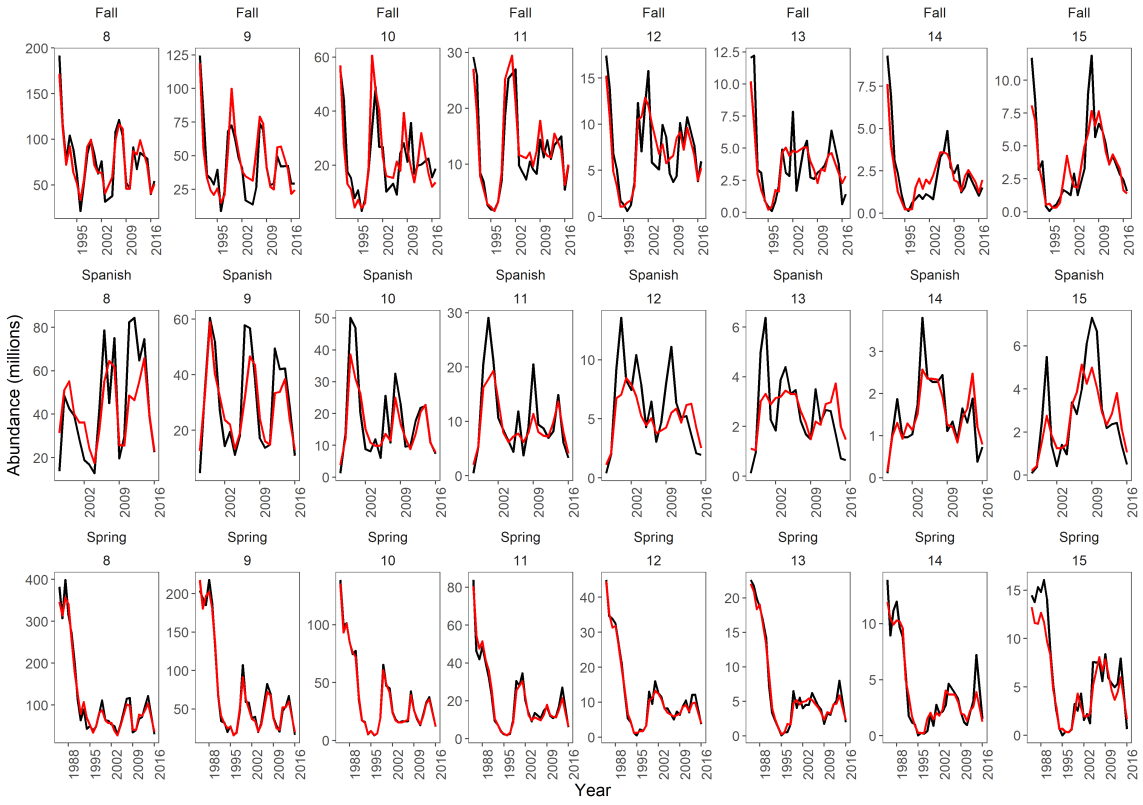


Figure F.21: EM2 standardized bubble plot of residuals at age for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative.

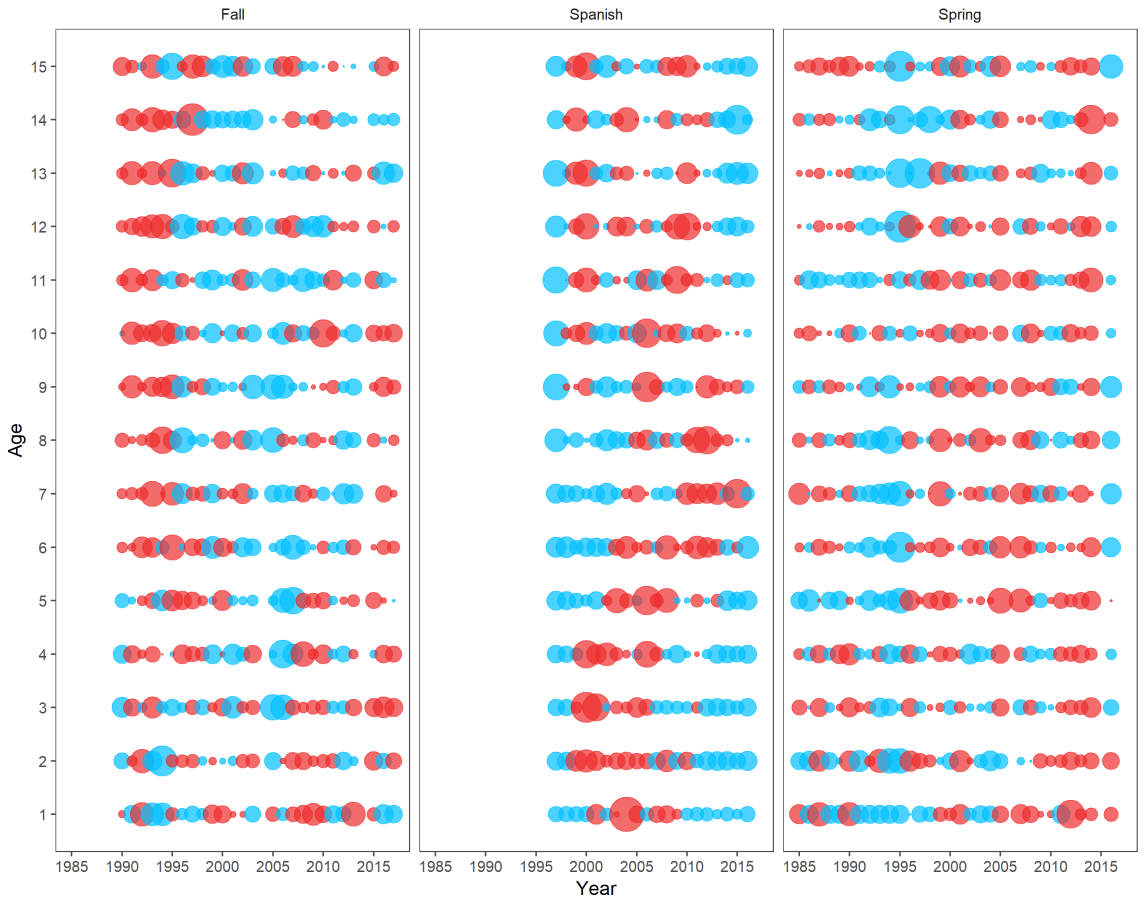


Figure F.22: EM2 standardized residuals for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO.

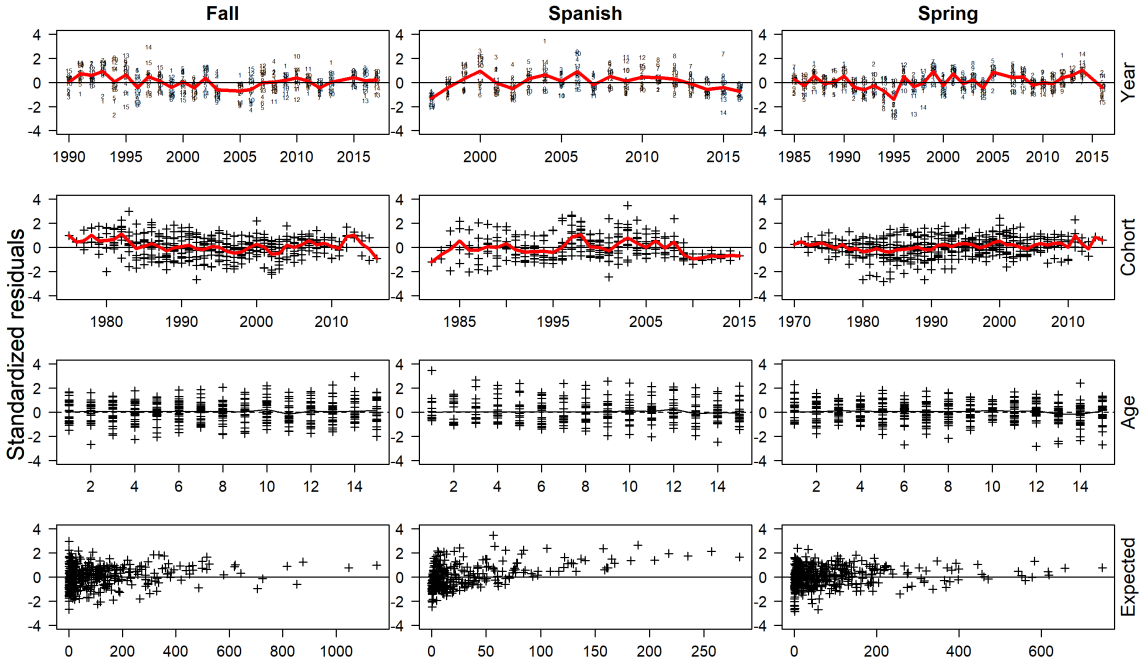


Figure F.23: EM2 estimated log catch numbers for ages 5-15+ (solid line) for American plaice in NAFO Divisions 3LNO, the shaded grey represents the region between the log lower catch bounds and the log upper catch bounds.

Figure F.24: EM2 predicted catch proportions at age (red) vs observed catch proportion at age (black) for American plaice in NAFO Divisions 3LNO.

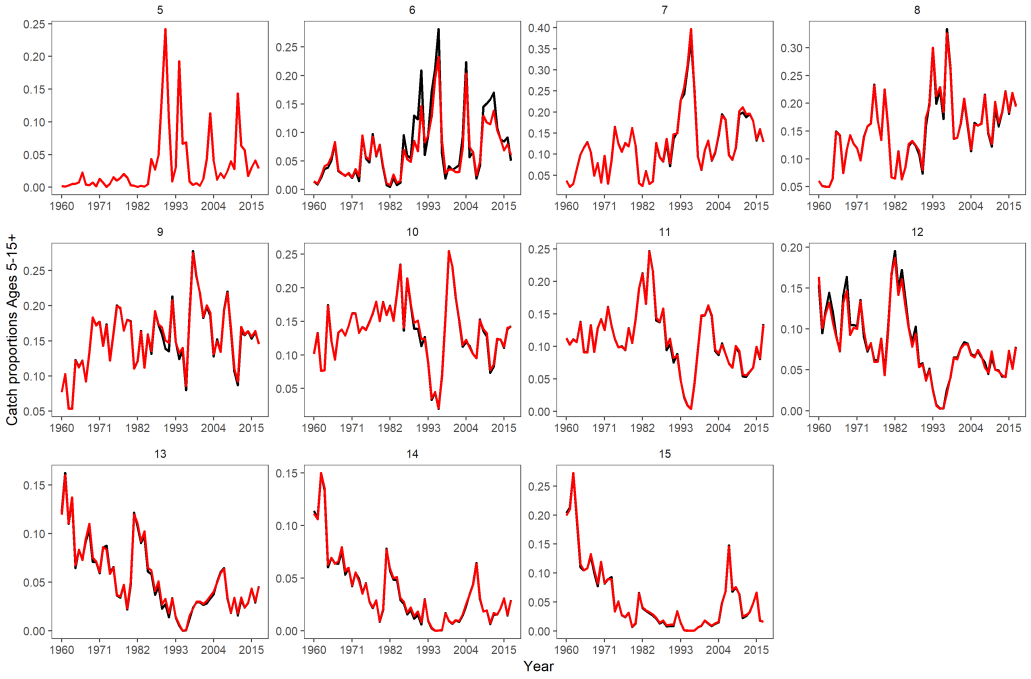


Figure F.25: EM2 standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO.

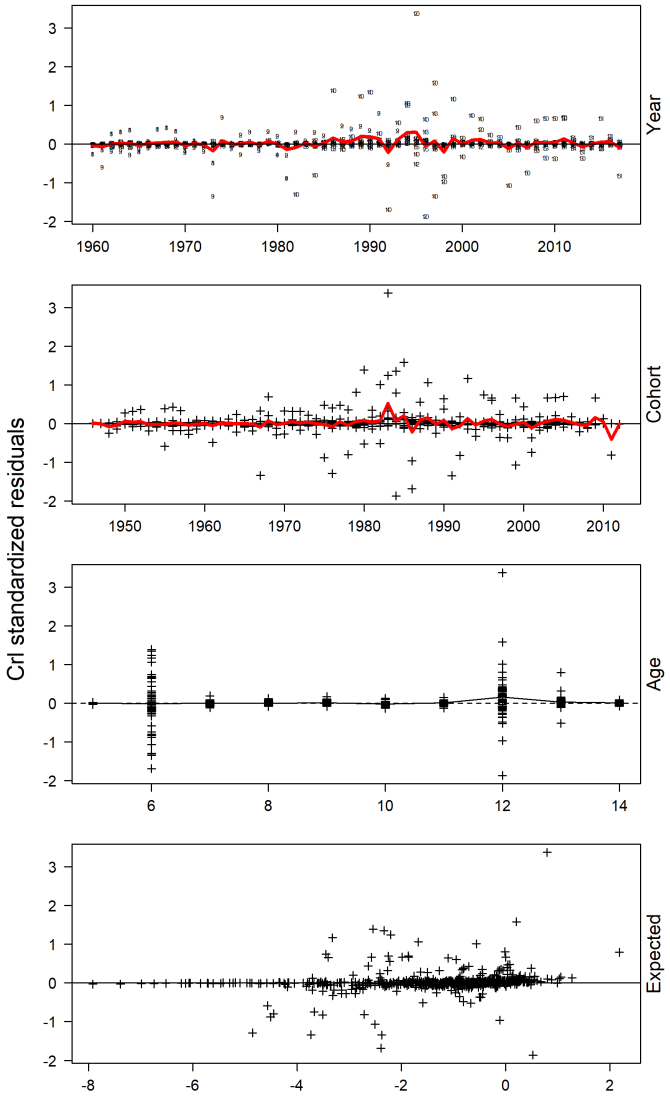


Figure F.26: EM2 bubble plot of standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO.

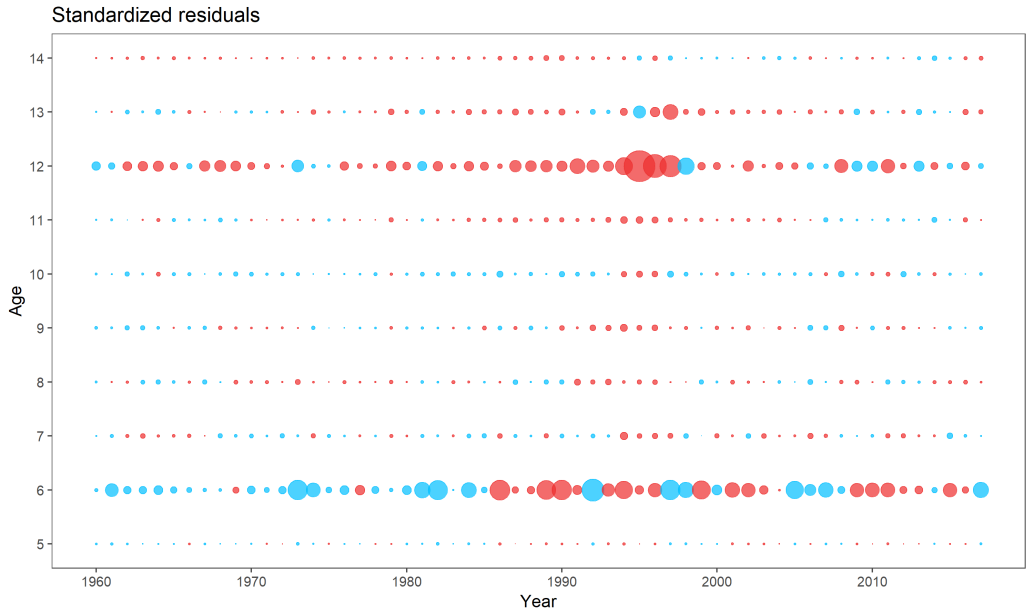


Figure F.27: EM2 predicted log fishing mortality rates for American plaice in NAFO Divisions 3LNO.

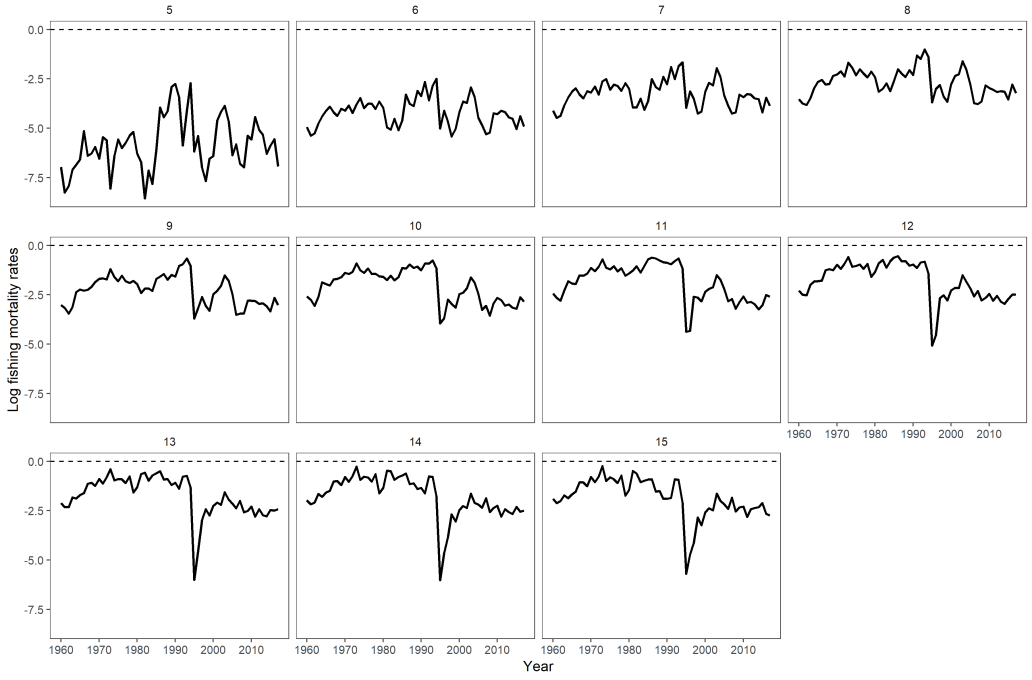


Figure F.28: EM2 predicted log fishing mortality deviations for American plaice in NAFO Divisions 3LNO.

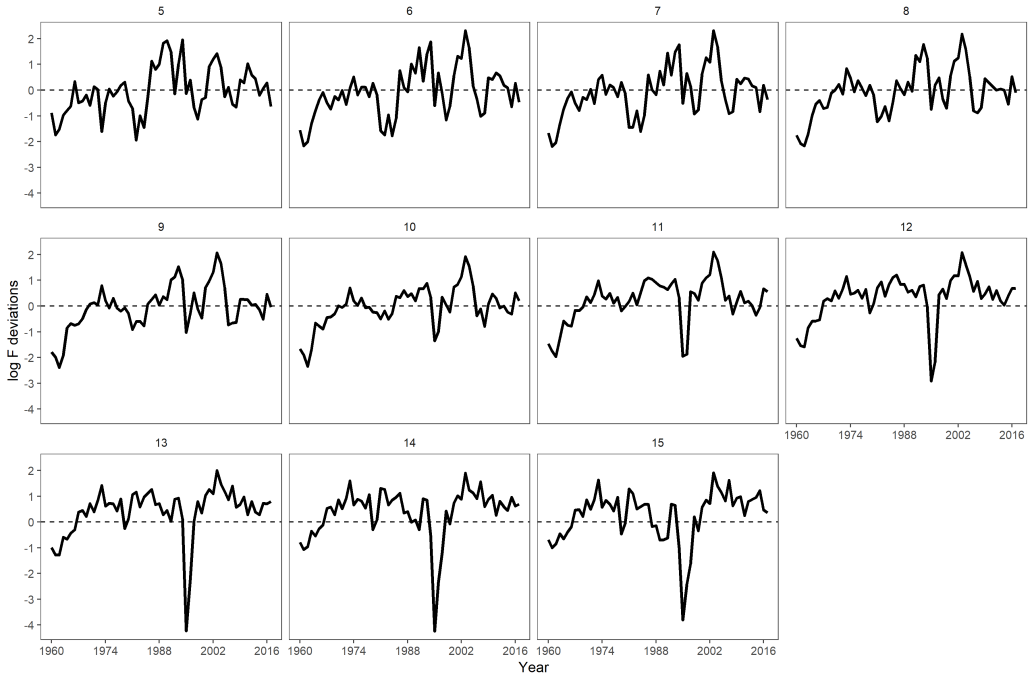


Figure F.29: EM2 predicted fishing mortality rates for American plaice in NAFO Divisions 3LNO.

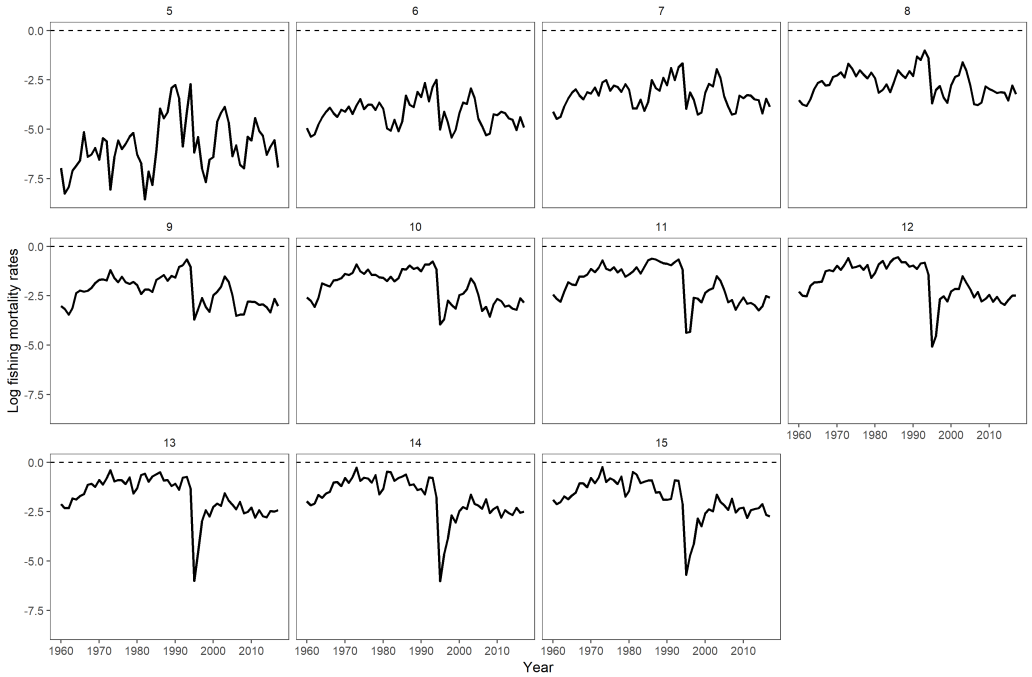


Figure F.30: EM2 predicted process errors at age for American plaice in NAFO Divisions 3LNO.

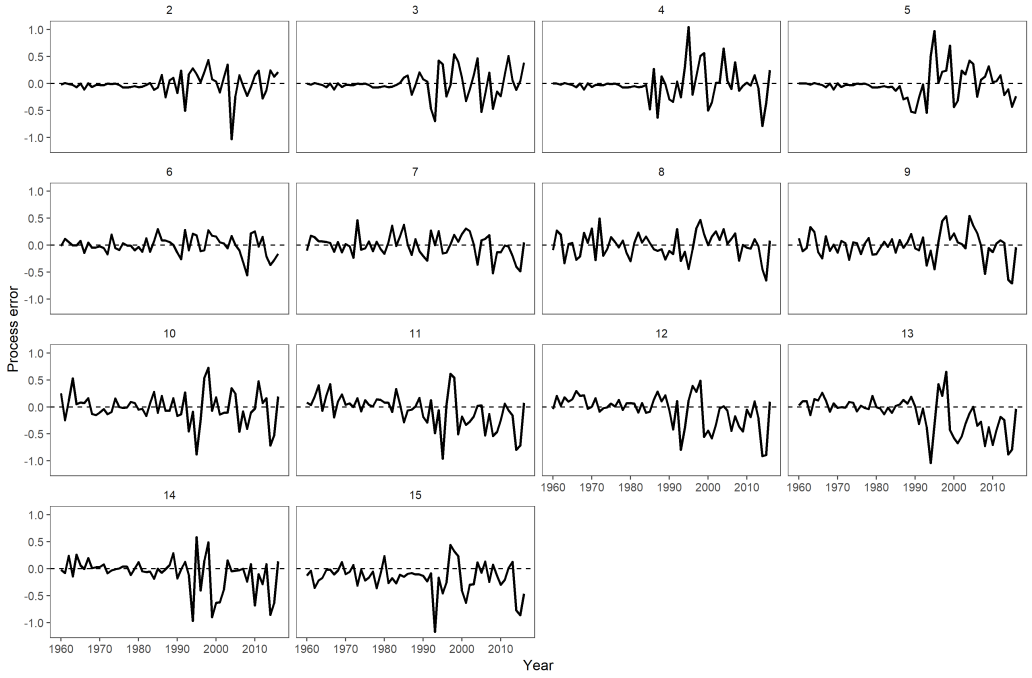


Figure F.31: EM2 bubble plot of predicted process errors at age for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative.

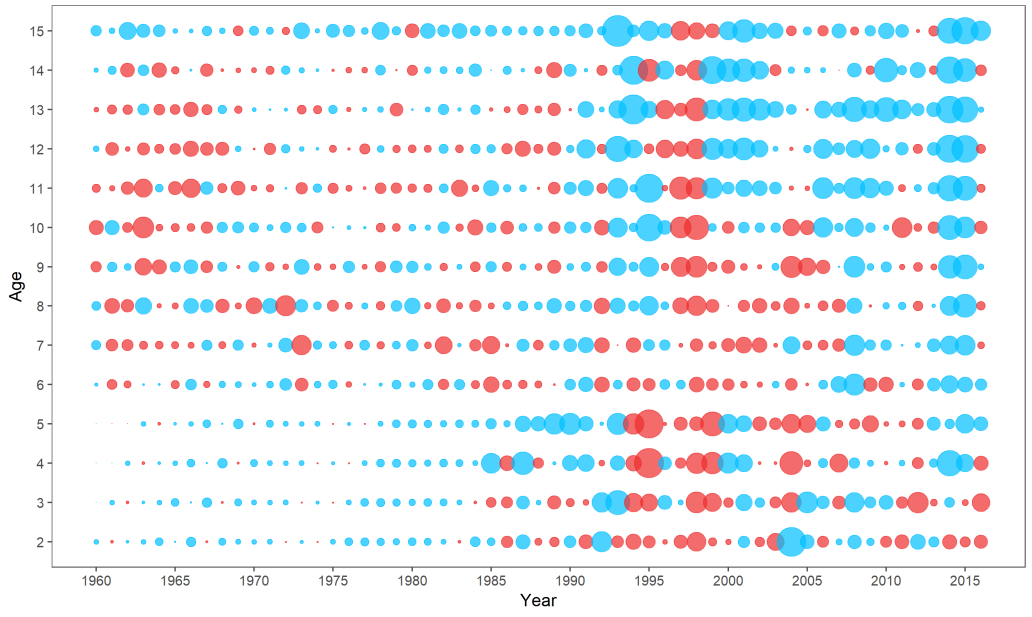


Figure F.32: EM2 survey catchability patterns for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. A separate catchability parameter is estimated for two gear types (Engel and Campelen) for the spring and fall surveys for ages 1-4.

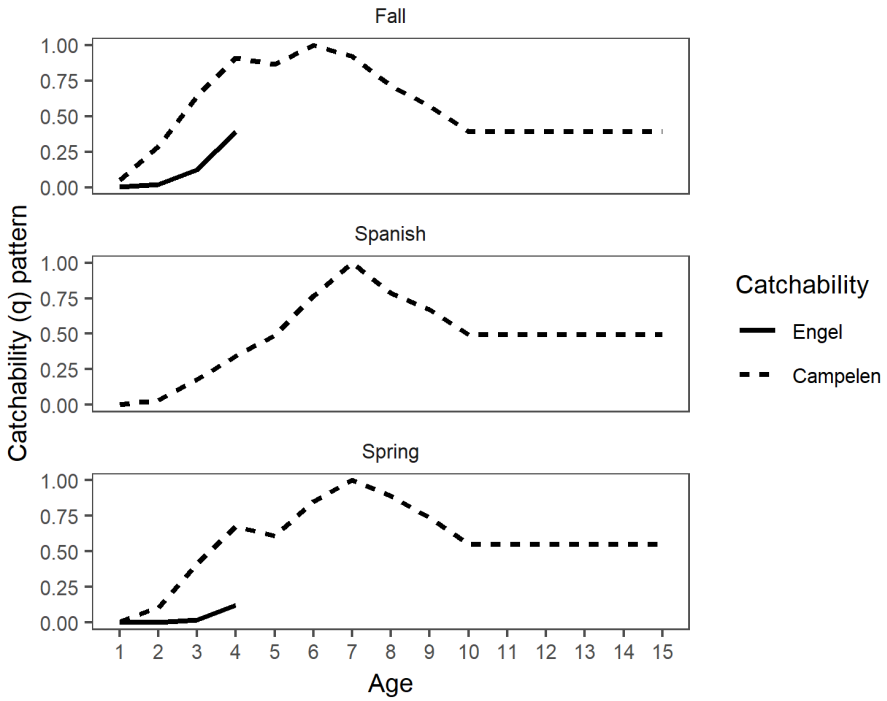


Figure F.33: EM2 (SSM) and VPA estimated spawning stock biomass for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent 95% confidence intervals.

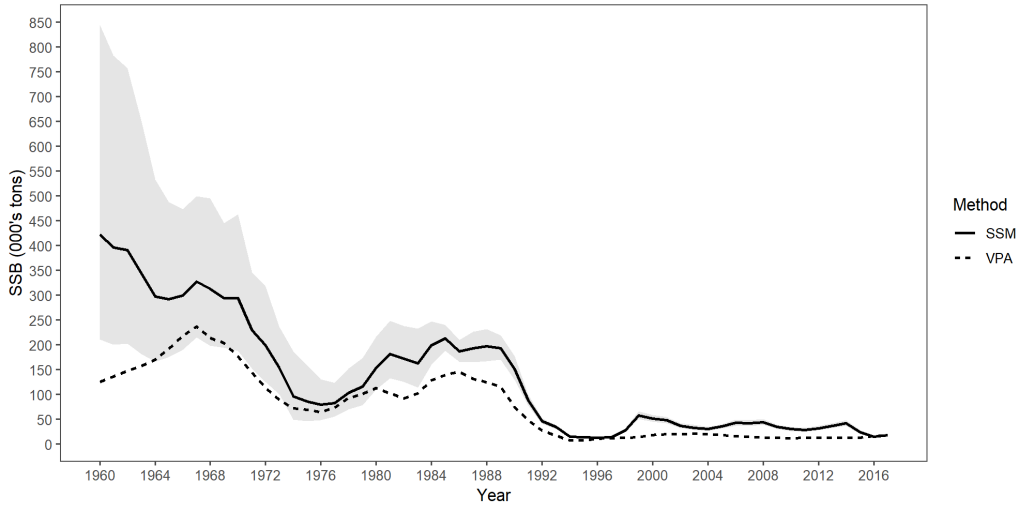


Figure F.34: EM2 (SSM) and VPA estimated average fishing mortality rates (ages 9-14) for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent 95% confidence intervals.

EM3

The process model is identical to the process model described in 3.2.1. The observation model differs from that described in 3.2.2 by: 1) no correlations across survey and crls observations; $N(0, cv_s I_{s,a,y})$, $LN(0, \sigma_C)$.

	EM3
nll	4858
nparms	75
AIC	9867
BIC	10245

Figure F.35: EM3 estimated population abundance, spawning stock biomass (SSB), average fishing mortality rate (ages 9-14) and recruitment for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent 95% confidence intervals.

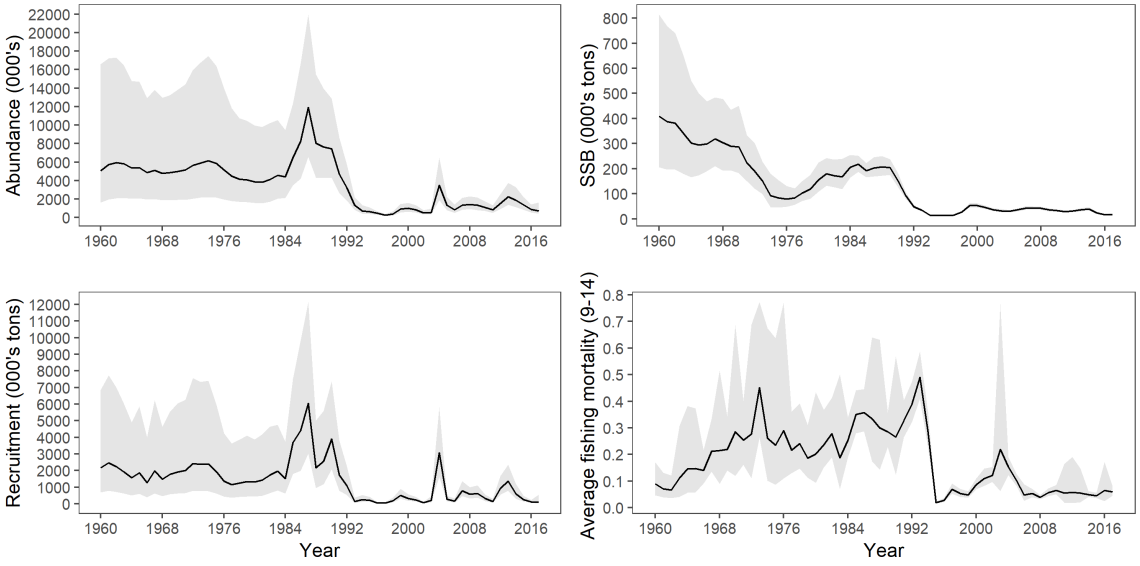


Figure F.36: EM3 predicted survey fits (red) and observed survey indices at age (black) for ages 1-7 for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO.

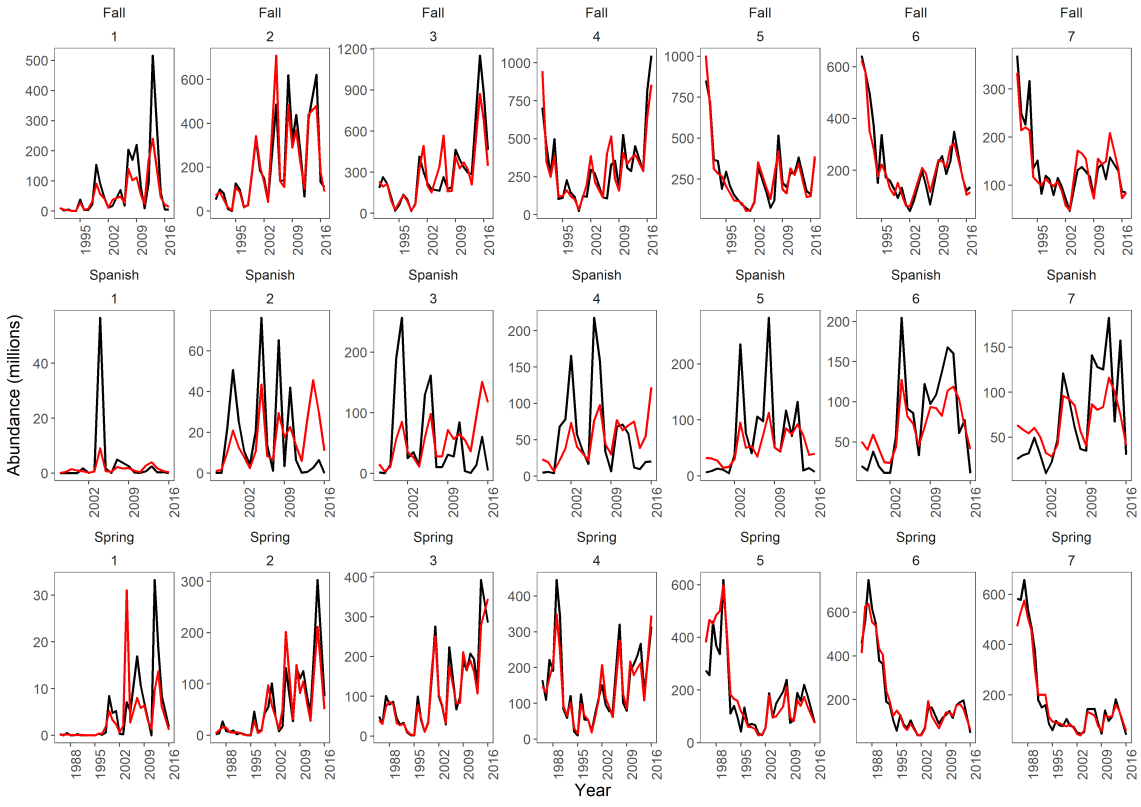


Figure F.37: EM3 predicted survey fits (red) and observed survey indices at age (black) for ages 8+ for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO.

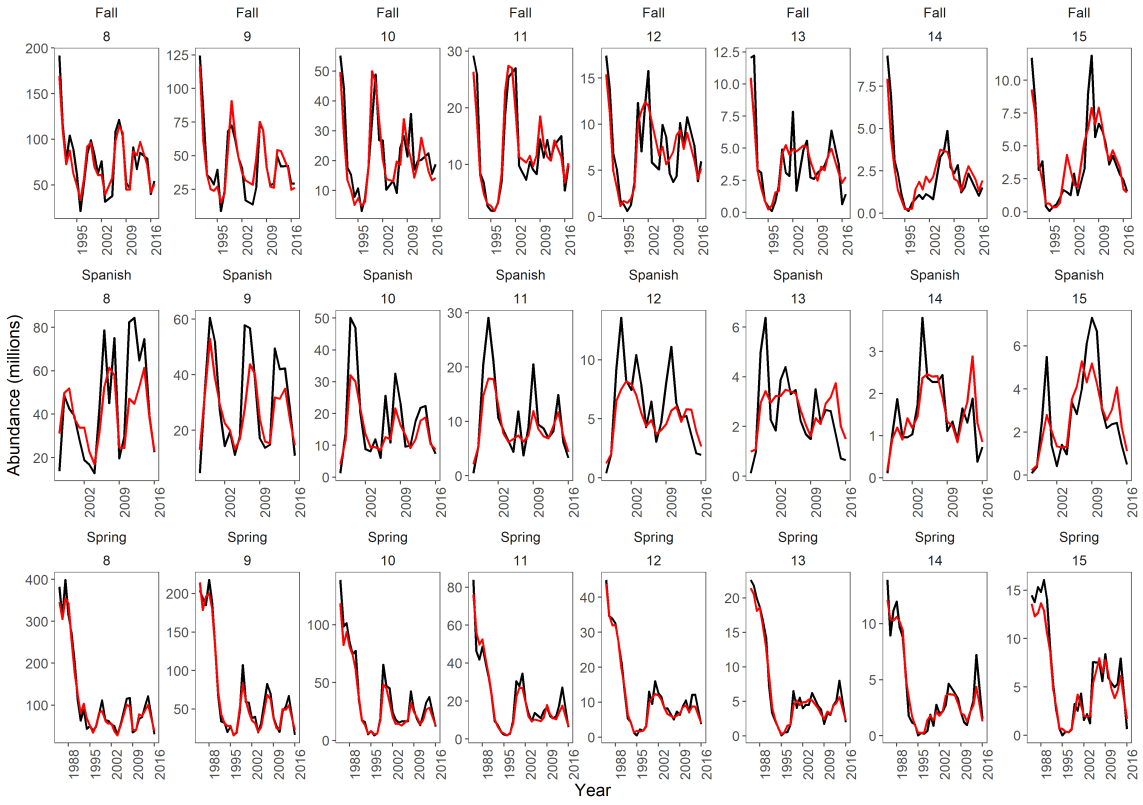


Figure F.38: EM3 standardized bubble plot of residuals at age for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative.

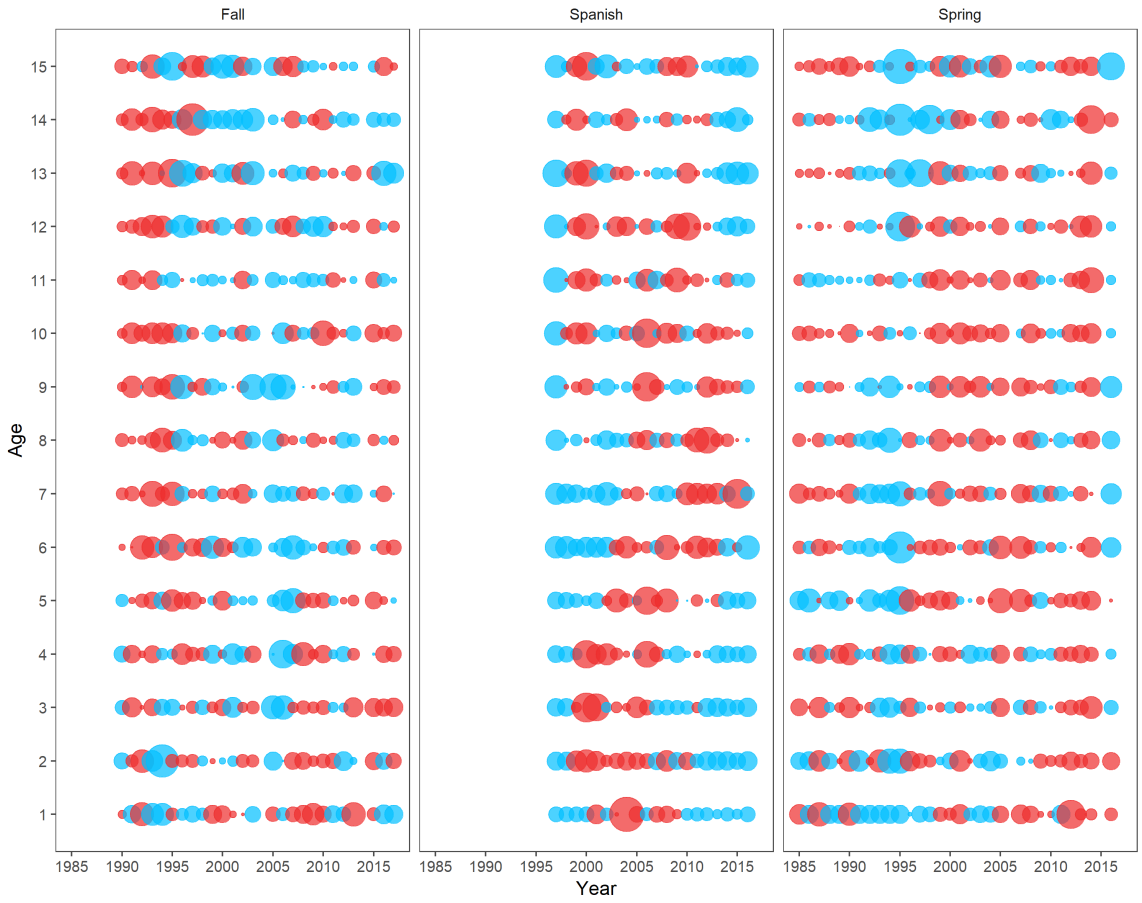


Figure F.39: EM3 standardized residuals for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO.

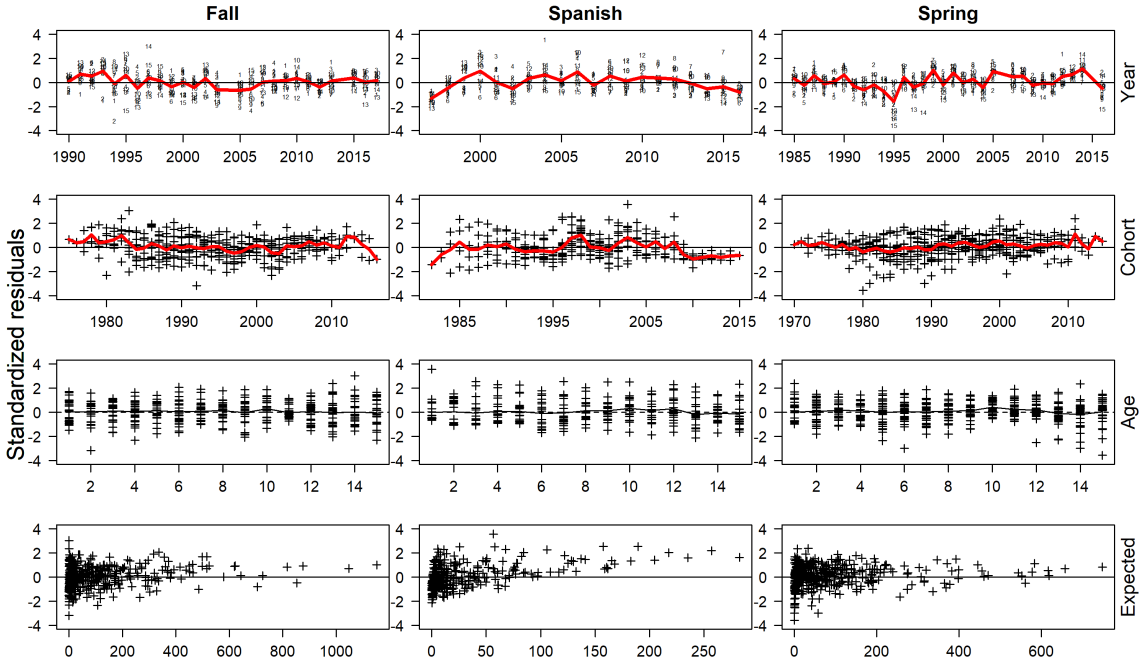


Figure F.40: EM3 estimated log catch numbers for ages 5-15+ (solid line) for American plaice in NAFO Divisions 3LNO, the shaded grey represents the region between the log lower catch bounds and the log upper catch bounds.

Figure F.41: EM3 predicted catch proportions at age (red) vs observed catch proportion at age (black) for American plaice in NAFO Divisions 3LNO.

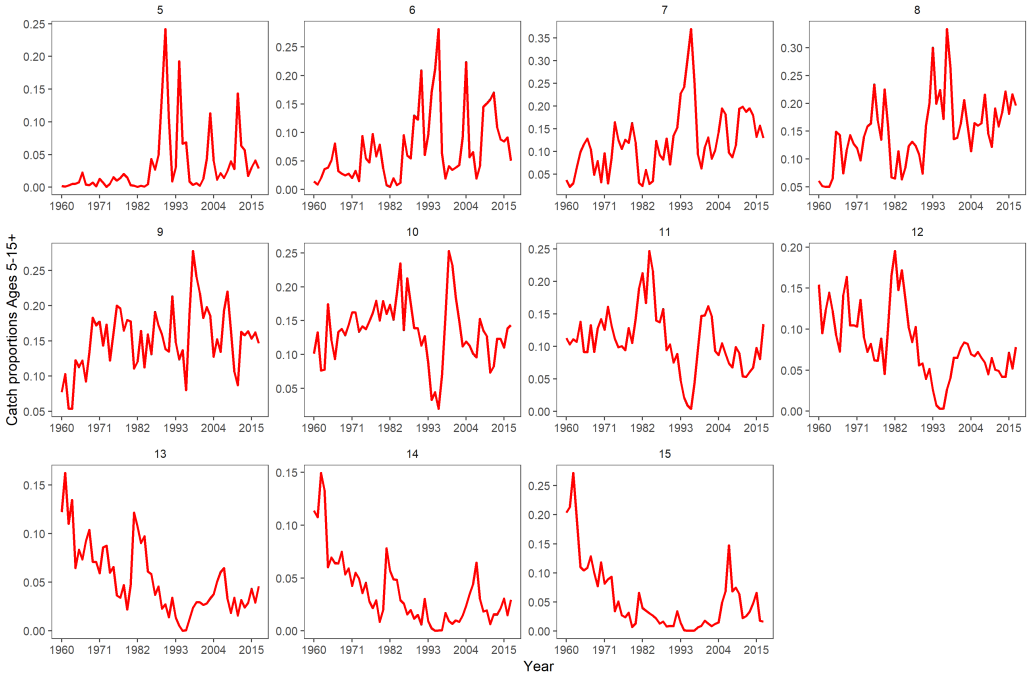


Figure F.42: EM3 standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO.

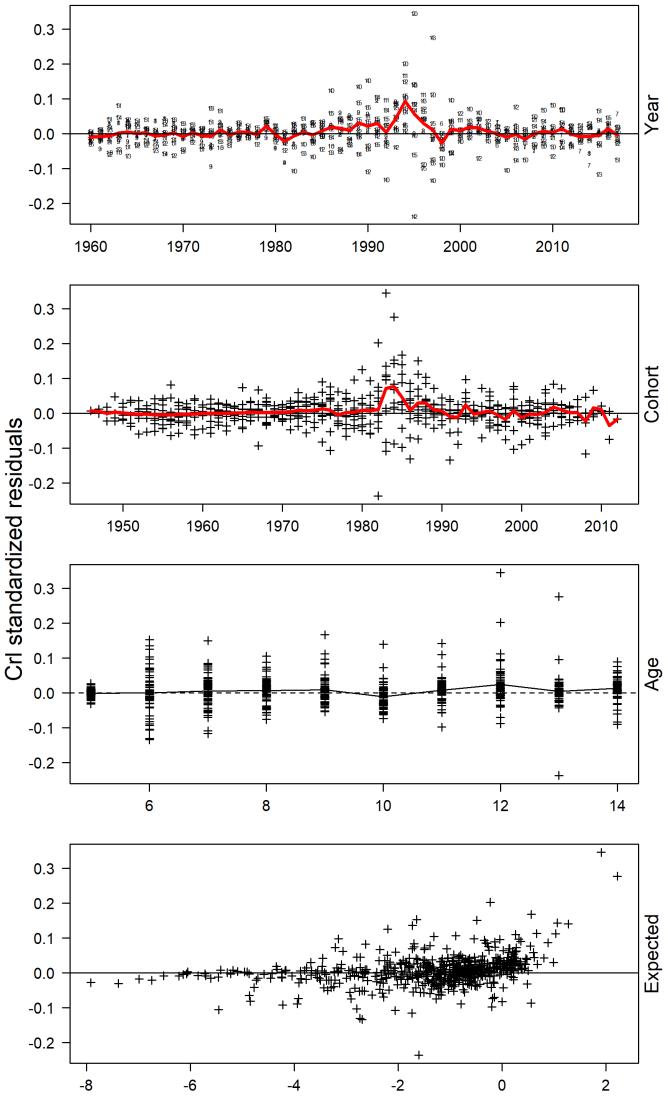


Figure F.43: EM3 bubble plot of standardized residuals for continuation ratio logits for American plaice in NAFO Divisions 3LNO.

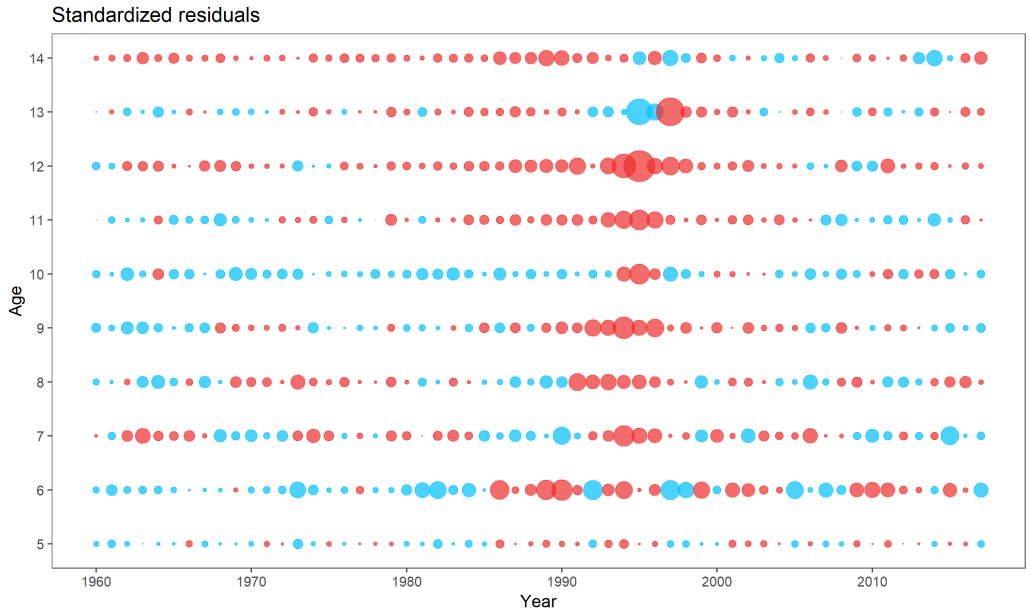


Figure F.44: EM3 predicted log fishing mortality rates for American plaice in NAFO Divisions 3LNO.

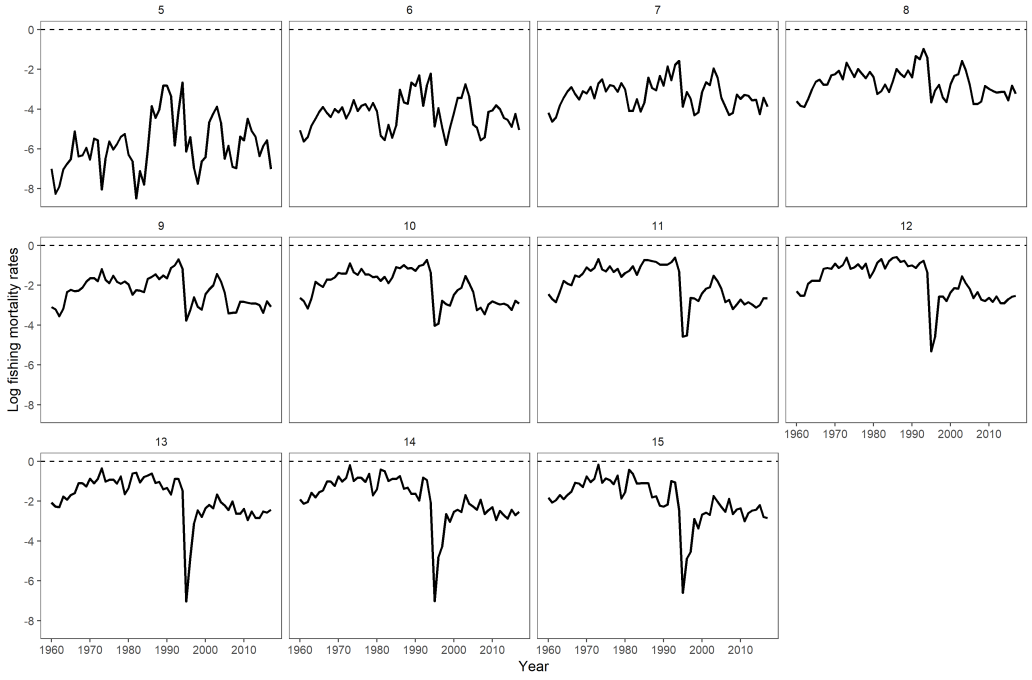


Figure F.45: EM3 predicted log fishing mortality deviations for American plaice in NAFO Divisions 3LNO.

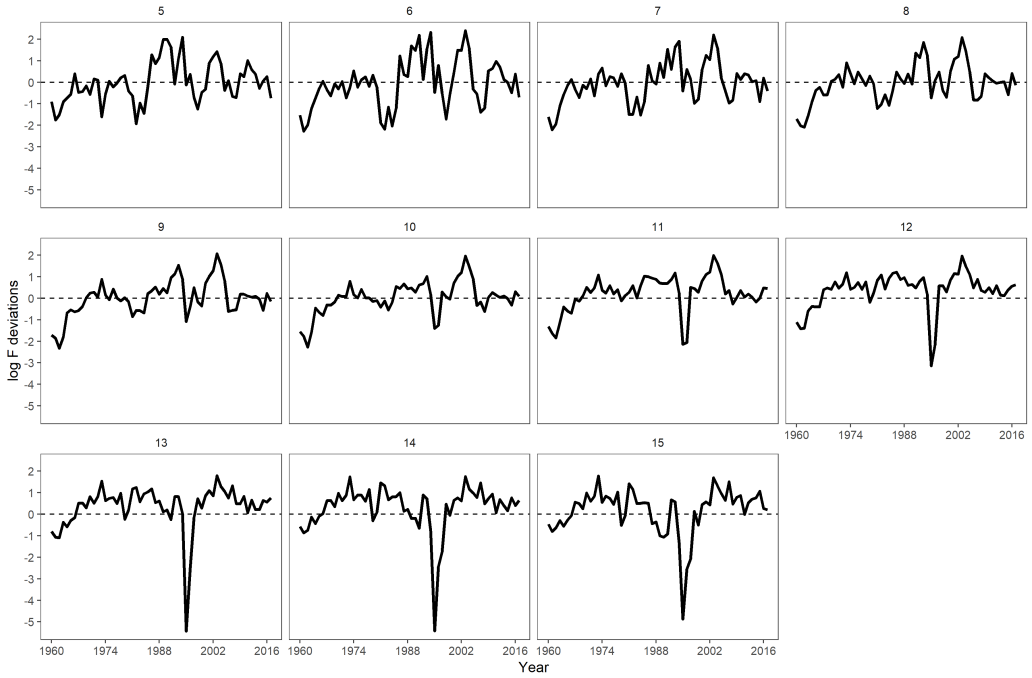


Figure F.46: EM3 predicted fishing mortality rates for American plaice in NAFO Divisions 3LNO.

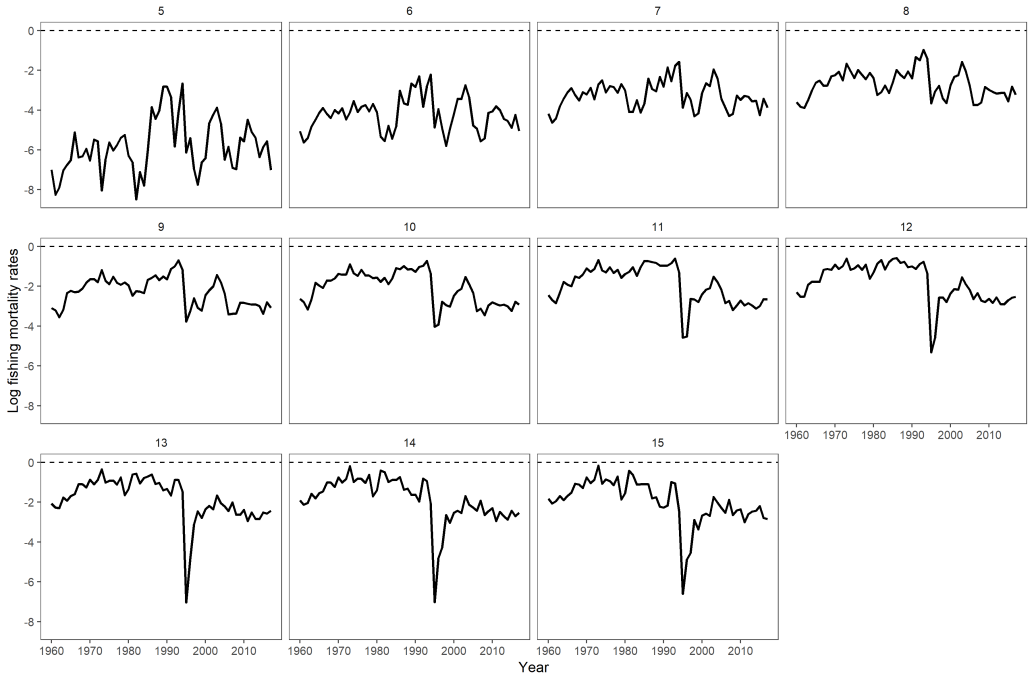


Figure F.47: EM3 predicted process errors at age for American plaice in NAFO Divisions 3LNO.

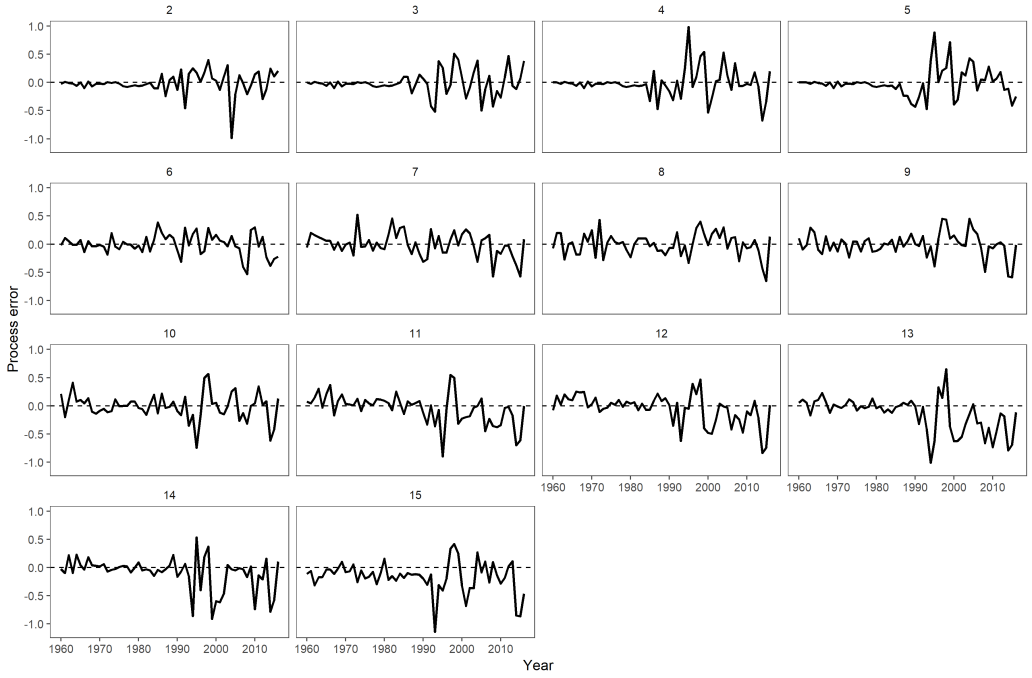


Figure F.48: EM3 bubble plot of predicted process errors at age for American plaice in NAFO Divisions 3LNO. Red is positive, blue is negative.

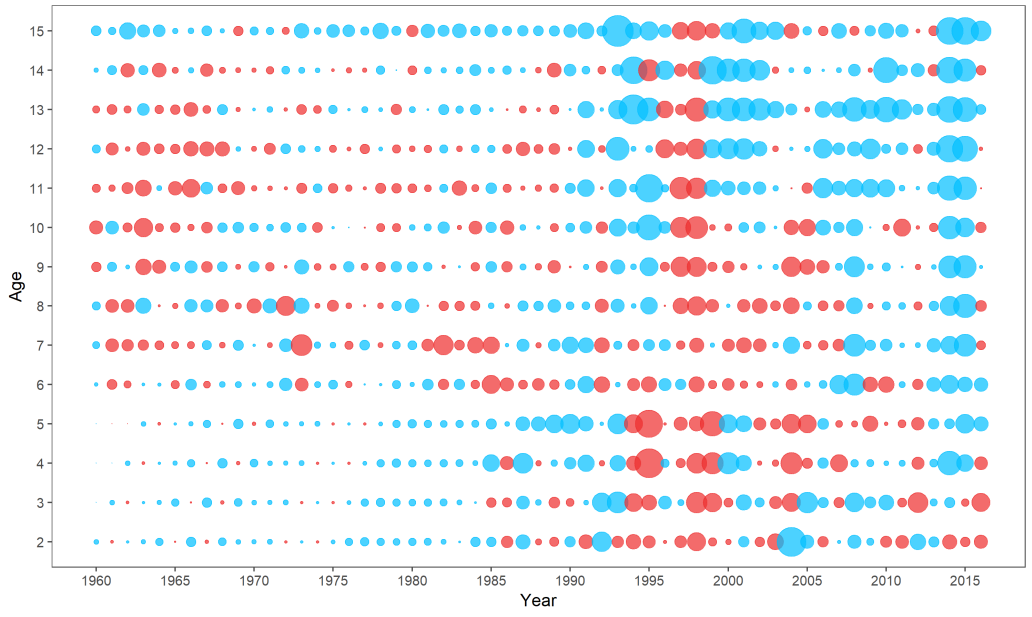


Figure F.49: EM3 survey catchability patterns for Canadian fall, spring and Spanish surveys for American plaice in NAFO Divisions 3LNO. A separate catchability parameter is estimated for two gear types (Engel and Campelen) for the spring and fall surveys for ages 1-4.

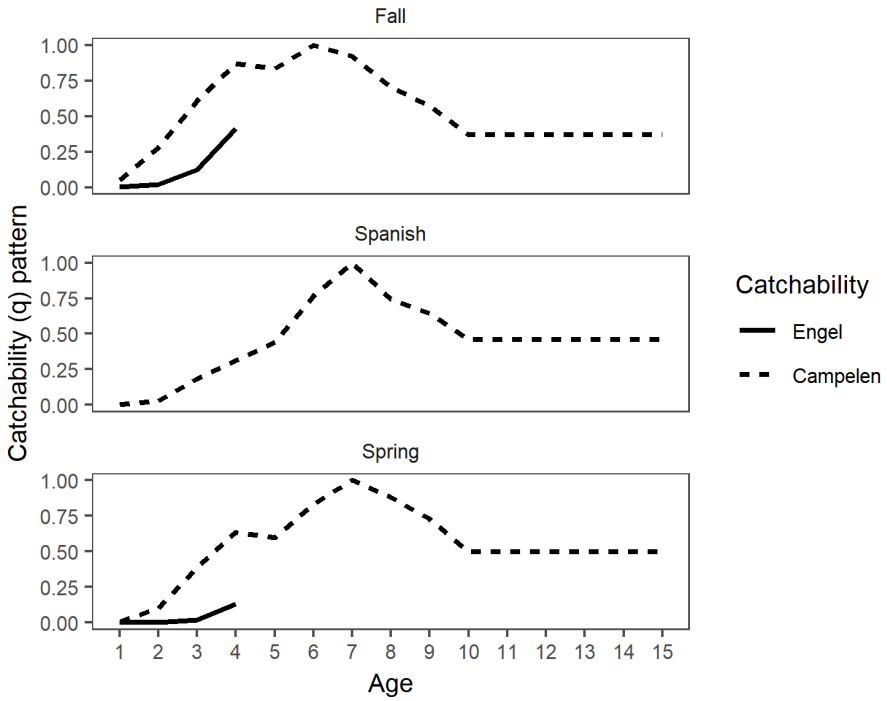


Figure F.50: EM3 (SSM) and VPA estimated spawning stock biomass for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent 95% confidence intervals.

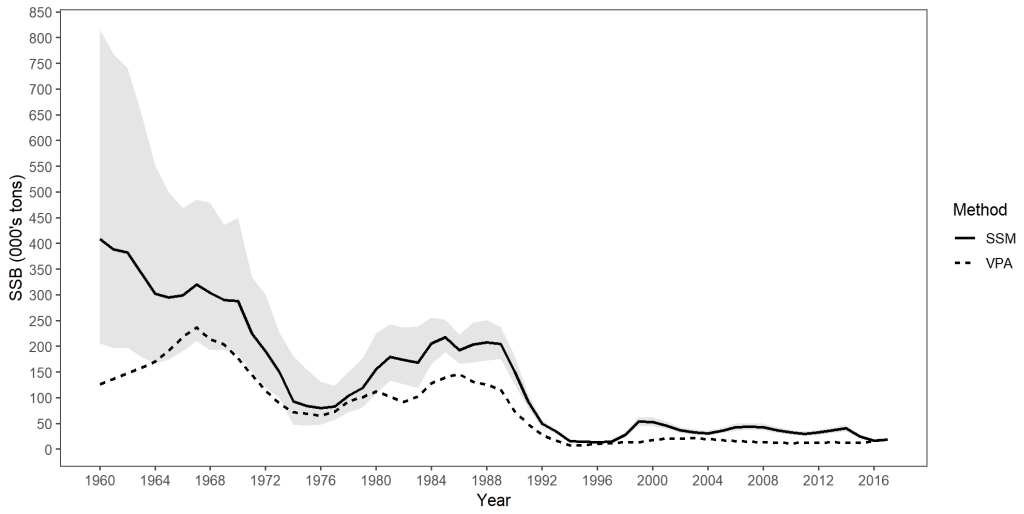
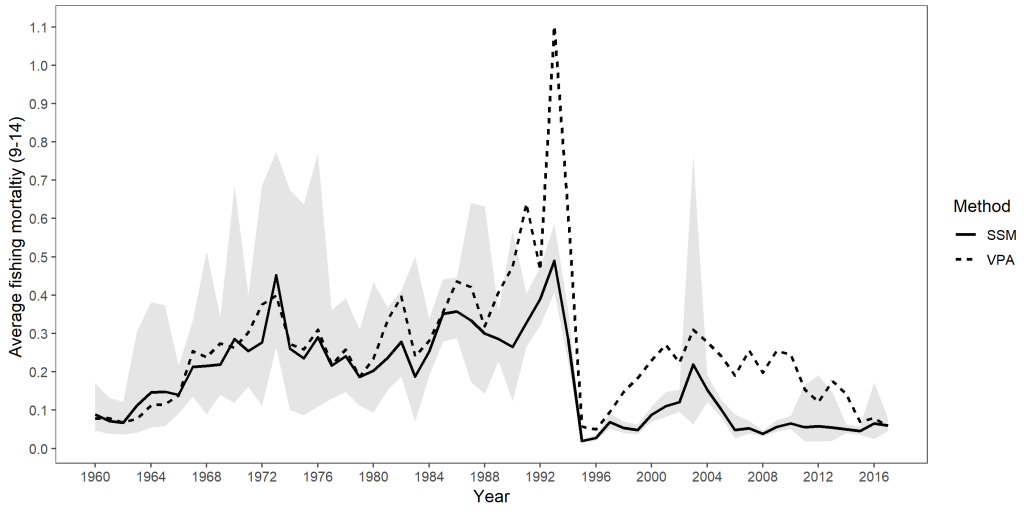


Figure F.51: EM3 (SSM) and VPA estimated average fishing mortality rates (ages 9-14) for years 1960-2017 for American plaice in NAFO Divisions 3LNO. The gray shaded regions represent 95% confidence intervals.



Appendix G

SSM supporting code

```
#data cleanup

#required libraries
library(reshape2)
library(ggplot2)
library(tidyr)
library(FSA)
library(stringr)
library(dplyr)

#define variables
end.year <- 2017
end.year.m1<-end.year-1
assess.year = 1960:end.year
age=1:15

#to read data into R
cnames = c("Year", "Landings")
landings = read.table("./Data/landings.txt", header = F, col.names = cnames)
landings = subset(landings,Year<=end.year)

cnames = c('Year',paste('Age',1:20,sep=""))
SRV.matrix  = read.table("Data/SpringSurvey.txt", header = F,
col.names =  cnames)
SRV.matrix = subset(SRV.matrix,Year<=end.year)

cnames = c('Year',paste('Age',0:20,sep=""))
FRV.matrix = read.table(file='Data/FallSurvey.txt',header=F,col.names=cnames)
FRV.matrix = subset(FRV.matrix,Year<=end.year)

cnames = c('Year',paste('Age',1:24,sep=""))
SS.matrix = read.table(file='Data/SpanishSurvey.txt',header=F,col.names=cnames)
SS.matrix = subset(SS.matrix,Year<=end.year)
```

```
cnames = c('Year',paste('Age',5:15,sep=""))
catch  = read.table(file='Data/catchage.txt',header=F,col.names=cnames)
catch = subset(catch,Year<=end.year)

cnames = c('Year',paste('Age',5:15,sep=""))
mat.all = read.table(file='Data/mat.txt',header=F,col.names=cnames)

#to read in catch weights
cnames = c('year',paste('age',5:15,sep=""))
catch_wt = read.table(file='Data/catch_wts.txt',header=F, col.names =cnames)
c_wt =  catch_wt[2:12]

#to read in Rajeev's weights
weights = read.csv("./Data/stock_weights.csv", header = TRUE)
weight = as.data.frame(weights)
weight$index = weight$springWt_3LNO/1000
nyear = c(1960:2017)
nages = c(1:21)

#to filter years of interest
stock_wt_mat = weight %>%  filter(year>1959, year<2018) %>%
select("year", "age","index") %>%
acast(year~age, value.var="index")

mean_sw_mat = stock_wt_mat

#to get mean value for first NAs
for(i in length(nages):1){
for(j in length(nyear):1){
if(is.na(mean_sw_mat[j,i])){mean_sw_mat[j,i]=mean(mean_sw_mat[(j+1):(j+3),i])
break}
}
}

#to set rest of NAs to mean value
for(i in length(nages):1){
for(j in length(nyear):1){
if(is.na(mean_sw_mat[j,i])){mean_sw_mat[j,i]=mean_sw_mat[j+1,i]}
}
```

```
}

#to get plus group
sw_plus_mat = mean_sw_mat
#sw_plus_mat[,15] = rowMeans(mean_sw_mat[,15:21])
sw_plus_mat = sw_plus_mat[,1:15]
colnames(sw_plus_mat) = c(paste('Age',1:15,sep=""))

#subset for assessment
FRV.matrix = cbind(FRV.matrix[,c(1,3:16)], rowSums(FRV.matrix[,17:22]))
colnames(FRV.matrix) = c('Year',paste('Age',1:15,sep=""))
SRV.matrix = cbind(SRV.matrix[,c(1,2:15)],rowSums(SRV.matrix[,16:21]))
colnames(SRV.matrix) = c('Year',paste('Age',1:15,sep=""))
SS.matrix = cbind(SS.matrix[,c(1,2:15)],rowSums(SS.matrix[,16:25]))
colnames(SS.matrix) = c('Year',paste('Age',1:15,sep=""))

mat = subset(mat.all,mat.all[,1] %in% assess.year)
mat = mat[,2:12]

year = catch[,1]
catch = catch[,2:12]

#to make vector data frames
vec_func = function(mdat){
vdat = melt(mdat,id=c("Year"))
vdat$Age = as.numeric(substr(vdat$variable,4,5))
vdat$variable=NULL
vdat$index=vdat$value
vdat$value=NULL
return(vdat)
}

FRV.vec = vec_func(FRV.matrix)
FRV.vec$survey = 'Fall'
SRV.vec = vec_func(SRV.matrix)
SRV.vec$survey = 'Spring'
SS.vec = vec_func(SS.matrix)
SS.vec$survey = "Spanish"

indices = rbind(FRV.vec,SRV.vec,SS.vec)

temp = cbind(year,catch)
colnames(temp) = c('Year',colnames(catch))
```

```
catch.vec = vec_func(temp)

#to convert to vectors
temp = cbind(nyear,c_wt)
colnames(temp) = c('Year',paste('age',5:15,sep=""))
cwt.vec = vec_func(temp)

temp = cbind.data.frame(nyear,sw_plus_mat)
colnames(temp) = c('Year',paste('age',1:15,sep=""))
stockwt.vec= vec_func(temp)

temp = cbind(year,mat)
colnames(temp) = c('Year',colnames(mat))
mat.vec = vec_func(temp)

#fs
indices$fs=NA
indices$fs[indices$survey=="Spring"] = 5.5/12
indices$fs[indices$survey=="Fall"] = 10.5/12
indices$fs[indices$survey=="Spanish"] = 5.5/12

iyear = as.numeric(factor(assess.year))-1
iage = as.numeric(factor(age))-1

indices$iyear = iyear[match(indices$Year,assess.year)]
indices$iage = iage[match(indices$Age,age)]
indices$isurvey = as.numeric(factor(indices$survey))-1

indices$surv_age = as.factor(paste(indices$survey,"_",str_pad(indices$Age , 2,
pad = "0"),sep=''))
indices$isd = as.numeric(indices$surv_age)-1

indices$surv_year = as.factor(paste(indices$survey,"_",indices$Year,sep=''))
indices$is_year = as.numeric(indices$surv_year)-1

#separate q for each survey and youngest ages (for trawl switch)
indices$qname = paste(indices$survey,":",indices$Age,sep='')
ind = indices$Age >=10
indices$qname[ind] = paste(indices$survey[ind],":10p",sep='')
ind = (indices$Age <= 4)&(indices$survey == 'Fall')&(indices$Year<1995)
indices$qname[ind] = paste("Fall_Engel:", indices$Age[ind])
ind = (indices$Age <= 4)&(indices$survey == 'Spring')&(indices$Year<=1995)
indices$qname[ind] = paste("Spring_Engel:", indices$Age[ind])
```

```
indices$iq = as.numeric(as.factor(indices$qname))-1

A = length(age)
Y = length(assess.year)
Ns = length(unique(indices$is_year))

M_matrix = matrix(0.2,nrow=Y,ncol=A,byrow=T)

#M = 0.53 for all ages for years 1989-1996
M_matrix[,1:3] = rep(0.50)
M_matrix[,4] = rep(0.30)
M_matrix[30:37,] = M_matrix[30:37,] + 0.33

olandings = landings$Landings/1000
catcht = rowSums(catch)

p_ya = t(catch/catcht)

pya_sum = apply(p_ya, 2, rcumsum)
pya_sum = pya_sum[1:10,]

pi_ya = p_ya[1:10,]/(pya_sum)

x_ya = log(pi_ya/(1-pi_ya))
crl  = t(x_ya)

#tmb formatted data
tmb.data = list(
M = M_matrix,
weight = as.matrix(sw_plus_mat),
mat = as.matrix(mat),
midy_weight = as.matrix(c_wt),
index = indices$index,
olandings = olandings,
iyear = indices$iyear,
iage = indices$iage,
isurvey = indices$isurvey,
isd = indices$isd ,
is_year = indices$is_year,
iq = indices$iq,
fs = indices$fs,
A = A,
```

```
Y = Y,
Ns = Ns, #total number of survey years
isurvey1 = unlist(tapply(indices$isurvey,indices$is_year,unique)),
crl =crl
)

#give names for surveys and q
names(tmb.data$iq) = indices$qname
names(tmb.data$isd) = indices$surv_age

#to get log index and index in tmb.data
tmb.data$index = as.numeric(tmb.data$index)

tmb.data$landings_L = tmb.data$olandings
# high uncertainty
tmb.data$landings_U = 2*tmb.data$olandings
# low uncertainty
ind = (landings$Year>=1977)&(landings$Year<=1982) | (landings$Year>=1994)
&(landings$Year<=2010)
tmb.data$landings_U[ind] = 1.2*tmb.data$olandings[ind]
# moderate uncertainty
ind = (landings$Year>=1983)&(landings$Year<=1993) | (landings$Year>=2011)
tmb.data$landings_U[ind] = 1.5*tmb.data$olandings[ind]


# tmb.data$log_landings = log(tmb.data$olandings)
tmb.data$log_landings = rep(0,length(log(tmb.data$olandings)))
tmb.data$log_landings_L = log(tmb.data$landings_L)
tmb.data$log_landings_U = log(tmb.data$landings_U)

tmb.data$std_log_landings = 0.05

save(tmb.data,indices, file='Data/tmb.RData')
save(mat,  catch, crl,catch.vec, mat.vec, stockwt.vec,cwt.vec, pi_ya, indices,
landings, M_matrix,p_ya, file = "Data/3LNO.Rdata" )

###to run model
library(Matrix)
library(TMB)

compile("./CPP/Fit1.cpp")
dyn.load(dynlib("./CPP/Fit1"))
```

```
load("./Data/3LNO.RData")
load("./Data/tmb.RData")

setparm <- function(opt, rep){

start.parameters <- parameters
uname <- names(parameters)
names.par <- names(opt$par)
for(i in 1:length(uname)){
ind <- names.par==uname[i]
start.parameters[[i]] = opt$par[ind]
}
start.parameters$log_F_devt = t(rep$log_F_dev)
return(start.parameters)
}


No = c(50000,10290,10290,10290,10290,10290, 13110, 22860, 14140, 11880, 5890,
4157,3390,1236,3078)

n.survey.sd=length(unique(tmb.data$isd))
n.survey=3
Y=tmb.data$Y
A=tmb.data$A

#to set parm start values and bounds
parameters <- list(
log_R=c(10,10),
log_qparm=rep(0,length(unique(tmb.data$iq))),
log_cv_index=rep(log(0.3),n.survey.sd),
log_std_log_R=log(1),
log_F_mean=matrix(log(0.05),nrow=Y,ncol=A-4),
log_std_log_F = rep(log(0.2),A-4),
log_std_pe = matrix(log(0.2),nrow=Y-1,ncol=A-1),
log_std_crl = matrix(log(0.2),nrow=Y,ncol=A-5),
logit_ar_index_age = rep(0,n.survey),
logit_ar_logF = rep(0,3),
logit_ar_crl = c(0,0),
logit_ar_pe = rep(-10,2),
logit_ar_logRec = 0,
log_F_devt=t(matrix(log(0.001),nrow=Y,ncol=A-4,byrow=T)),
log_Nt=t(matrix(log(No),nrow=Y,ncol=A,byrow=T))
)
```

```
parameters.L <- list(
log_R=c(-10,-10),
log_qparm=rep(-Inf,length(unique(tmb.data$iq))),
log_cv_index=rep(log(0.001),n.survey.sd),
log_std_log_R=-10,
log_F_mean=matrix(-Inf,nrow=Y,ncol=A-4),
log_std_log_F = rep(log(0.001),A-4),
log_std_pe = matrix(log(0.001),nrow=Y-1,ncol=A-1),
log_std_crl = matrix(-10,nrow=Y,ncol=A-5),
logit_ar_index_age = rep(-10,n.survey),
logit_ar_logF = rep(-10,3),
logit_ar_crl = rep(-10,2),
logit_ar_pe = rep(-10,2),
logit_ar_logRec = -10
)

parameters.U <- list(
log_R=c(Inf,Inf),
log_qparm=rep(10,length(unique(tmb.data$iq))),
log_cv_index=rep(log(5),n.survey.sd),
log_std_log_R=Inf,
log_F_mean=matrix(Inf,nrow=Y,ncol=A-4),
log_std_log_F = rep(Inf,A-4),
log_std_pe = matrix(5,nrow=Y-1,ncol=A-1),
log_std_crl = matrix(5,nrow=Y,ncol=A-5),
logit_ar_index_age = rep(log(0.99/0.01),n.survey),
logit_ar_logF = c(log(0.99/0.01),log(0.99/0.01),log(0.99/0.01)),
logit_ar_crl = c(log(0.99/0.01),log(0.99/0.01)),
logit_ar_pe = c(log(0.99/0.01),log(0.99/0.01)),
logit_ar_logRec = log(0.99/0.01)
)


lower = unlist(parameters.L);
upper = unlist(parameters.U);

#parameter maps
mapF = matrix(NA,nrow=Y,ncol=A-4)
catch.year=landings$Year
agec = c(5:9,rep('10+',length = A-9))
#mapF = matrix(paste('M_',agec,sep=''),nrow=length(catch.year),ncol=A-4,byrow=T)
```

```
ind = catch.year<1995
mapF[ind,] = matrix(paste('F_',agec,'_pre_95',sep=''),
nrow=length(catch.year[ind]),ncol=A-4,byrow=T)
ind = catch.year>=1995
mapF[ind,] = matrix(paste('F_',agec,'_post_95',sep=''),
nrow=length(catch.year[ind]),ncol=A-4,byrow=T)

rownames(mapF) =catch.year

agec = c(1,rep('2-5',4),rep('6-9',4),rep('10-15',6))
n1 = paste('Fall_',agec,sep="")
agec = c(1,2,rep('3-5',3),rep('6-9',4),rep('10-15',6))
n3 = paste('Sprg_',agec,sep="")
agec = c(1,rep('2-5',4),rep('6-9',4),rep('10-15',6))
n2 = paste('Span_',agec,sep="")

map_index_cv = c(n1,n2,n3)

agec = c(5,rep('6+',length = A-6))
map_log_std_crl = matrix(paste('std_crl_',agec,sep=''),nrow=Y,ncol=A-5,byrow=T)

map_std_pe = matrix('std_pe',nrow=Y-1,ncol=A-1)

map = list(
#  logit_ar_index_age = factor(rep(NA,n.survey)),
logit_ar_logF = factor(c('age','year1','year1+')),
#  logit_ar_crl = factor(c(NA,NA)),
logit_ar_pe = factor(c(NA,NA)),
#  logit_ar_pe  = factor(c('EQ','EQ')),
#  logit_ar_pe  = factor(c(NA,NA)),
log_F_mean = factor(mapF),
log_std_log_F = factor(c("5",rep("6+",A-5))),
log_cv_index = factor(map_index_cv),
log_std_crl = factor(map_log_std_crl),
log_std_pe = factor(map_std_pe)
)

tp=parameters.L;
#tp$logit_ar_index_age = NULL
tp$logit_ar_logF = rep(-10,3);
#tp$logit_ar_crl = NULL;
tp$logit_ar_pe = NULL;
tp$log_F_mean=rep(-Inf,length(unique(as.vector(mapF))));
```

```
tp$log_std_log_F=rep(-Inf,length(unique(map$log_std_log_F)));
tp$log_cv_index=rep(-Inf,length(unique(map$log_cv_index)));
tp$log_std_crl=rep(log(0.02),length(unique(map$log_std_crl)));
tp$log_std_pe=rep(log(0.001),length(unique(map$log_std_pe)));
#tp$log_std_pe=NULL;
lower = unlist(tp);

tp=parameters.U;
#tp$logit_ar_index_age = NULL
tp$logit_ar_logF = rep(10,3);
#tp$logit_ar_crl = NULL;
tp$logit_ar_pe = NULL;
tp$log_F_mean=rep(Inf,length(unique(as.vector(mapF))));
tp$log_std_log_F=rep(log(3),length(unique(map$log_std_log_F)));
tp$log_cv_index=rep(log(5),length(unique(map$log_cv_index)));
tp$log_std_crl=rep(log(1),length(unique(map$log_std_crl)));
tp$log_std_pe=rep(log(10),length(unique(map$log_std_pe)));
#tp$log_std_pe=NULL;
upper = unlist(tp);

tmb.data$log_lowerM = log(tmb.data$landings_L)
tmb.data$log_upperM = log(tmb.data$landings_U)
tmb.data$use_pe <- 1;
tmb.data$std_log_landings= 0.02

obj <- MakeADFun(tmb.data,parameters,map=map,
random=c("log_F_devt","log_Nt"),
DLL = "Fit1",
control = list(trace=10,eval.max=2000,iter.max=300))

length(lower)
length(upper)
length(obj$par)

obj$gr(obj$par)

opt<-nlminb(obj$par,obj$fn,obj$gr,
upper=upper,lower=lower,
control = list(trace=10,eval.max=2000,iter.max=200))

opt<-nlminb(opt$par,obj$fn,obj$gr,
upper=upper,lower=lower,
control = list(trace=10,eval.max=2000,iter.max=200))
```

```
opt<-nlminb(opt$par,obj$fn,obj$gr,
upper=upper,lower=lower,
control = list(trace=10,eval.max=2000,iter.max=300))

rep = obj$report()

sd.rep <- sdreport(obj)

### code to get sd report including getting covariances of residuals

compile("./CPP/cov1.cpp")
dyn.load(dynlib("./CPP/cov1"))

sp = setparm(opt,rep)
sp$log_std_log_F = log(rep$std_log_F)
sp$log_F_mean = rep$log_F_mean
sp$log_cv_index <- log(rep$cv_index)
sp$logit_ar_index_age = rep$logit_ar_index_age
sp$logit_ar_logF <- rep$logit_ar_logF
sp$logit_ar_crl <- rep$logit_ar_crl
sp$logit_ar_pe <- rep$logit_ar_pe
sp$log_std_pe = log(rep$std_pe)
sp$log_Nt = t(rep$log_N)
sp$log_std_crl = log(rep$std_crl)
sp$resid_index = rep(0,length=length(tmb.data$index))
sp$resid_crl = matrix(0,nrow=tmb.data$Y,ncol=tmb.data$A-5)

map1 <- list(
log_R=factor(c(NA,NA)),
log_qparm=factor(rep(NA,length(unique(tmb.data$iq)))),
log_cv_index=factor(rep(NA,n.survey.sd)),
log_std_log_R=factor(NA),
log_F_mean=factor(matrix(NA,nrow=Y,ncol=A-4)),
log_std_log_F = factor(rep(NA,A-4)),
log_std_pe = factor(matrix(NA,nrow=Y-1,ncol=A-1)),
log_std_crl = factor(matrix(NA,nrow=Y,ncol=A-5)),
logit_ar_index_age = factor(rep(NA,n.survey)),
logit_ar_logF = factor(rep(NA,3)),
logit_ar_crl = factor(c(NA,NA)),
logit_ar_pe = factor(c(NA,NA)),
logit_ar_logRec = factor(NA),
log_F_devt=factor(t(matrix(NA,nrow=Y,ncol=A-4,byrow=T))),
```

```
log_Nt=factor(t(matrix(No*NA,nrow=Y,ncol=A,byrow=T)))
)

obj1 <- MakeADFun(tmb.data,sp,map=map1,DLL = "cov1")

rep1 = obj1$report();
sd.rep1<-sdreport(obj1)

ind=names(sd.rep1$value)=="resid_index"

resid.cov = sd.rep1$cov[ind,ind]
ch.cov = chol(resid.cov)

rep$index_Zresid = qr.solve(t(ch.cov), rep$resid_index)

# qqplot(qqnorm(rep$index_Zresid))
# abline(a=0,b=1,col='red')

## stuff just to check;
# ind1 = tmb.data$is_year==46
# covi = resid.cov[ind1,ind1]
# sdi = sqrt(diag(covi))
# corri = diag(1/sdi)%*%covi%*%diag(1/sdi)

ind=names(sd.rep1$value)=="resid_crl"

resid.cov = sd.rep1$cov[ind,ind]
ch.cov = chol(resid.cov)

crl_Zresid = qr.solve(t(ch.cov), as.vector(rep$resid_crl))

rep$crl_Zresid = matrix(crl_Zresid,nrow=tmb.data$Y,ncol=tmb.data$A-5)

# qqplot(qqnorm(crl_Zresid))
# abline(a=0,b=1,col='red')
#
# ## stuff just to check;
# covi = resid.cov
# sdi = sqrt(diag(covi))
# corri = diag(1/sdi)%*%covi%*%diag(1/sdi)
#
# matrix(corri[1,],nrow=tmb.data$Y,ncol=tmb.data$A-5)
#
```

```
# #}
save.image( file='Data/Fit1.RData')


###C++ code passed to TMB
#include <TMB.hpp>
#include "pnorm4.hpp" //Atomic functions for censored likelihoods
#include <iostream>

template <class Type>
Type trans(Type x){
return Type(2)*invlogit(x)-Type(1);
}

template<class Type>
Type cens(Type obs,Type ex, Type sd, Type upper){
return log(pnorm((log(upper)+(obs-ex))/sd)-pnorm((obs-ex)/sd));
}
template<class Type>
Type objective_function<Type>::operator() ()
{

//input data;
DATA_MATRIX(M);
DATA_MATRIX(weight);
DATA_MATRIX(mat);
DATA_MATRIX(midy_weight);
DATA_VECTOR(index);
DATA_VECTOR(olandings);
DATA_IVECTOR(iyear);
DATA_IVECTOR(iage);
DATA_IVECTOR(isurvey);
DATA_IVECTOR(isd);
DATA_IVECTOR(is_year);
DATA_IVECTOR(iq);
DATA_VECTOR(fs);
DATA_INTEGER(A);
DATA_INTEGER(Y);
DATA_INTEGER(Ns);
DATA_ARRAY(crl);
DATA_IVECTOR(isurvey1);
DATA_VECTOR(landings_L);
DATA_VECTOR(landings_U);
```

```
DATA_VECTOR(log_landings);
DATA_VECTOR(log_landings_L);
DATA_VECTOR(log_landings_U);
DATA_SCALAR(std_log_landings);
DATA_VECTOR(log_lowerM);
DATA_VECTOR(log_upperM);
DATA_INTEGER(use_pe);

int n = index.size();
Type one = 1.0;
Type zero = 0.0;

//define fixed parameters;
PARAMETER_VECTOR(log_R);
PARAMETER_VECTOR(log_qparm);
PARAMETER_VECTOR(log_cv_index);
PARAMETER(log_std_log_R);
PARAMETER_MATRIX(log_F_mean);
PARAMETER_VECTOR(log_std_log_F);
PARAMETER_ARRAY(log_std_pe);
PARAMETER_ARRAY(log_std_crl);
PARAMETER_VECTOR(logit_ar_index_age);
PARAMETER_VECTOR(logit_ar_logF);
PARAMETER_VECTOR(logit_ar_crl);
PARAMETER_VECTOR(logit_ar_pe);
PARAMETER(logit_ar_logRec);

//define random effects
PARAMETER_ARRAY(log_F_devt);
PARAMETER_ARRAY(log_Nt);

array<Type> log_F_dev = log_F_devt.transpose();
array<Type> log_N = log_Nt.transpose();

//  set bounds on parameters
Type std_log_R = exp(log_std_log_R);
vector <Type> std_log_F = exp(log_std_log_F);
vector<Type> cv_index = exp(log_cv_index);
//  Type std_pe = exp(log_std_pe);

//  Type std_crl = exp(log_std_crl);
Type ar_logF_age = invlogit(logit_ar_logF(0));
Type ar_logF1_year = invlogit(logit_ar_logF(1));
```

```
Type ar_logF_year = invlogit(logit_ar_logF(2));
Type ar_crl_age = invlogit(logit_ar_crl(0));
Type ar_crl_year = invlogit(logit_ar_crl(1));
Type ar_pe_age = invlogit(logit_ar_pe(0));
Type ar_pe_year = invlogit(logit_ar_pe(1));
Type ar_logRec = invlogit(logit_ar_logRec);
vector<Type> ar_index_age = exp(logit_ar_index_age)/
(one + exp(logit_ar_index_age));

//containers
matrix<Type> N(Y,A);
array<Type> pe(Y-1,A-1);
vector<Type> log_Rec_dev(Y);
vector<Type> log_Rec(Y);

matrix<Type> EC(Y,A-4);
matrix<Type> ECW(Y,A-4);

vector<Type> C_tot(Y);
vector<Type> CW_tot(Y);
vector<Type> log_landings_pred(Y);
vector<Type> std_landings_resid(Y);
vector<Type> landings_resid(Y);

vector<Type> Elog_index(n);
vector<Type> Eindex(n);
vector<Type> resid_index(n);
vector<Type> std_resid_index(n);
vector<Type> log_q_vec = log_qparm(iq);
vector<Type> resid_rec(n);

matrix<Type> p_ya(A-4,Y);
matrix<Type> pya_sum(A-4,Y);
matrix<Type> pi_ya(A-4,Y);
matrix<Type> pred_crl(Y,A-5);
matrix<Type> resid_crl(Y,A-5);
matrix<Type> std_resid_crl(Y,A-5);
array<Type> std_crl(Y,A-5);
array<Type> std_pe(Y-1,A-1);

//SD report objects
matrix<Type> B_matrix(Y,A);
matrix<Type> SSB_matrix(Y,A-4);
```

```
vector<Type> biomass(Y);
vector<Type> log_biomass(Y);
vector<Type> ssb(Y);
vector<Type> log_ssb(Y);
vector<Type> aveF_914(Y);
vector<Type> log_aveF_914(Y);


//initialize the negative log likelihood
Type nll = zero;
using namespace density;

//to model F (no fishing up to age 5)
matrix<Type> F(Y,A-4);
matrix<Type> Z(Y,A);
matrix<Type> log_F = log(F.array());

int i,j;
for(i = 0;i < Y;++i){
for(j = 0;j < A-5;++j){std_crl(i,j) = exp(log_std_crl(i,j));}
}
for(i = 0;i < Y-1;++i){
for(j = 0;j < A-1;++j){std_pe(i,j) = exp(log_std_pe(i,j));}
}
for(i = 0;i < Y;++i){
for(j = 0;j < A-4;++j){
log_F(i,j) = std_log_F(j)*log_F_dev(i,j) + log_F_mean(i,j);
F(i,j) = exp(log_F(i,j));
}
}

for(int i = 0;i < Y;++i){
for(int j = 0;j < A;++j){
if(j<4){Z(i,j) = M(i,j);}
if(j>=4){Z(i,j)= F(i,j-4)+M(i,j);}
}}


//compute process errors
for(int i = 1;i < Y;++i){
for(int j = 1;j < A-1;++j){
pe(i-1,j-1) = log_N(i,j) - log_N(i-1,j-1) + Z(i-1,j-1);
}
```

```
int j=A-1;
pe(i-1,j-1) = log_N(i,j) - log(exp(log_N(i-1,j-1)-Z(i-1,j-1))+
exp(log_N(i-1,j)-Z(i-1,j)));
}

//calculate recruitment deviations;
for(int i = 0;i < Y;++i){
if(i<=32){log_Rec_dev(i)= log_N(i,0) - log_R(0);}
if(i>32){log_Rec_dev(i)= log_N(i,0) - log_R(1);}
log_Rec(i) = log_N(i,0);
}

for(int i = 0;i < Y;++i){
for(int j = 0;j < A;++j){
N(i,j)=exp(log_N(i,j));}}


//Baranov catch equation predictions and residuals
for(int i = 0;i < Y;++i){
C_tot(i) = zero;
CW_tot(i) = zero;
for(int j = 0;j < A-4;++j){
EC(i,j) = N(i,j+4)*((one - exp(-one*Z(i,j+4)))*F(i,j)/Z(i,j+4));
ECW(i,j) = EC(i,j)*midy_weight(i,j);
C_tot(i) += EC(i,j);
CW_tot(i) += ECW(i,j);
} }

log_landings_pred = log(CW_tot);
landings_resid = log_landings - log_landings_pred;
std_landings_resid = landings_resid/std_log_landings;

//age composition catch
for(int i = 0;i < Y;++i){
for(int j = 0;j < A-4;++j){
p_ya(j,i) = EC(i,j)/C_tot(i);
}}

Type total;
for(int i = 0;i <Y;++i){
total=zero;
pya_sum(0,i) = one;
for(int j = 1;j < A-4;++j){
```

```
total+=p_ya(j-1,i);
pya_sum(j,i) = (one-total);
}}

for(int i = 0;i <Y;++i){
for(int j = 0;j < A-4;++j){
pi_ya(j,i) = p_ya(j,i)/pya_sum(j,i);}}


for(int i = 0;i<Y;i++){
for(int j = 0;j<A-5;j++){
pred_crl(i,j)= log(pi_ya(j,i)/(one - pi_ya(j,i)));
resid_crl(i,j) = crl(i,j) - pred_crl(i,j);
std_resid_crl(i,j) = resid_crl(i,j)/std_crl(i,j);}}

//Survey index predictions, and residuals;

vector<Type> std_index_vec(n);
matrix<Type> mresid_index(Ns,A);
matrix<Type> msd_index(Ns,A);

int ia,iy,iy1,is;
for(i = 0;i < n;++i){
ia = iage(i);
iy = iyear(i);
is = isd(i);
iy1 = is_year(i);

Elog_index(i) = log_q_vec(i) + log_N(iy,ia) - fs(i)*Z(iy,ia);
Eindex(i) = exp(Elog_index(i));
std_index_vec(i) = cv_index(is)*Eindex(i);
resid_index(i) = index(i) - Eindex(i);
std_resid_index(i) = resid_index(i)/std_index_vec(i);

mresid_index(iy1,ia) = resid_index(i);
msd_index(iy1,ia) = std_index_vec(i);
}

//NEGATIVE LOGLIKELIHOODS
//Index OBSERVATION MODEL

vector<Type> del(A);
vector<Type> sd_del(A);
```

```
for(i = 0;i < Ns;++i){
iy = isurvey1(i);
del = vector<Type>(mresid_index.row(i));
sd_del = vector<Type>(msd_index.row(i));
nll += VECSCALE(AR1(ar_index_age(iy)),sd_del)(del);
}

//Landings censored nll;
for(int i = 0;i < Y;++i){
nll-=censored_bounds(log_landings(i),log_landings_pred(i),std_log_landings,
-log_lowerM(i),log_upperM(i));
}


array<Type> temp(Y,A-5);
temp = resid_crl.array();
nll += VECSCALE(SEPARABLE(AR1(ar_crl_age),AR1(ar_crl_year)),std_crl)(temp);

//PROCESS MODEL

//Log recruitS;
nll += SCALE(AR1(ar_logRec),std_log_R)(log_Rec_dev);


//Log F
//yearxage correlation on second:last ages, RW first age;

vector<Type> delF1(35);
vector<Type> delF2(Y-35);
for(int j = 0;j < 35;++j){delF1(j) = log_F_dev(j,0);}
for(int j = 35;j < Y;++j){delF2(j-35) = log_F_dev(j,0);}

nll += AR1(ar_logF1_year)(delF1);
nll += AR1(ar_logF1_year)(delF2);

array<Type> log_F1_dev1(35,A-5);
array<Type> log_F1_dev2(Y-35,A-5);
for(int i = 1;i < A-4;++i){
for(int j = 0;j < 35;++j){log_F1_dev1(j,i-1) = log_F_dev(j,i);}
for(int j = 35;j < Y;++j){log_F1_dev2(j-35,i-1) = log_F_dev(j,i);}
}
```

```
nll += SEPARABLE(AR1(ar_logF_age),AR1(ar_logF_year))(log_F1_dev1);
nll += SEPARABLE(AR1(ar_logF_age),AR1(ar_logF_year))(log_F1_dev2);

//Process error

if(use_pe==1){
nll += VECSCALE(SEPARABLE(AR1(ar_pe_age),AR1(ar_pe_year)),std_pe)(pe);
}

//Useful output
//Biomass and SSB
for(int i=0;i<Y;++i){
for(int j=0;j<A;++j){
B_matrix(i,j) = weight(i,j)*N(i,j);
}}

for(int i=0;i<Y;i++){
for(int j=0;j<A-4;++j){
SSB_matrix(i,j) = mat(i,j)*B_matrix(i,j+4);
}}

for(int i = 0;i < Y;++i){
biomass(i) = zero;
ssb(i) = zero;
for(int j = 0;j < A;++j){
biomass(i) += B_matrix(i,j);
if(j>=4){ssb(i) += SSB_matrix(i,j-4);}
}}

log_biomass = log(biomass);
log_ssb = log(ssb);

//pop size weighted ave F(9-14)
Type tni;

for(int i = 0;i < Y;++i){
aveF_914(i) = zero;
tni = zero;
for(int j =8 ;j < 14;++j){
aveF_914(i) += F(i,j-4)*N(i,j);
tni += N(i,j);
}
aveF_914(i) = aveF_914(i)/tni;
```

```
}

log_aveF_914 = log(aveF_914);

REPORT(N);
REPORT(F);
REPORT(Z);
REPORT(B_matrix);
REPORT(SSB_matrix);
REPORT(biomass);
REPORT(ssb);
REPORT(aveF_914);
REPORT(log_Rec);
REPORT(C_tot);
REPORT(pred_crl);
REPORT(resid_crl);
REPORT(std_resid_crl);
REPORT(p_ya);
REPORT(pya_sum);
REPORT(pi_ya);
REPORT(CW_tot);
REPORT(EC);
REPORT(ECW);
REPORT(landings_resid);
REPORT(log_landings_pred);
REPORT(std_landings_resid);
REPORT(std_log_F);
REPORT(std_pe);
REPORT(std_crl);
REPORT(cv_index);
REPORT(std_index_vec);

REPORT(ar_crl_age);
REPORT(ar_crl_year);
REPORT(ar_logF_age);
REPORT(ar_logF_year);
REPORT(ar_index_age);
REPORT(ar_logRec);
REPORT(ar_pe_year);
REPORT(ar_pe_age);

REPORT(logit_ar_index_age);
REPORT(logit_ar_logF);
```

```
REPORT(logit_ar_crl);
REPORT(logit_ar_pe);
REPORT(logit_ar_logRec);

REPORT(Elog_index);
REPORT(Eindex);
REPORT(resid_index);
REPORT(std_resid_index);
REPORT(mresid_index);
REPORT(msd_index);

REPORT(log_N);
REPORT(log_Rec_dev);
REPORT(log_R);
REPORT(log_F_mean);
REPORT(log_F_dev);
REPORT(log_F1_dev1);
REPORT(log_F1_dev2);
REPORT(delF1);
REPORT(delF2);
REPORT(log_F);
REPORT(pe);
REPORT(log_qparm);

ADREPORT(log_landings_pred);
ADREPORT(log_biomass);
ADREPORT(log_ssb);
ADREPORT(log_aveF_914);
ADREPORT(log_Rec);
ADREPORT(log_N);
ADREPORT(ECW);
ADREPORT(log_qparm);

ADREPORT(log_qparm);

ADREPORT(std_log_F);
ADREPORT(std_pe);
ADREPORT(std_log_R);
ADREPORT(cv_index);
ADREPORT(ar_logF_age);
ADREPORT(ar_logF1_year);
ADREPORT(ar_logF_year);
ADREPORT(ar_pe_year);
```

```
ADREPORT(ar_pe_age);
ADREPORT(ar_logRec);
ADREPORT(ar_index_age);
ADREPORT(ar_crl_year);
ADREPORT(ar_crl_age);
ADREPORT(ssb);
ADREPORT(aveF_914);



return nll;
}


###C++ code for residuals
#include <TMB.hpp>
#include "pnorm4.hpp" //Atomic functions for censored likelihoods
#include <iostream>

template <class Type>
Type trans(Type x){
return Type(2)*invlogit(x)-Type(1);
}

template<class Type>
Type objective_function<Type>::operator() ()
{

//input data;
DATA_MATRIX(M);
DATA_MATRIX(weight);
DATA_MATRIX(mat);
DATA_MATRIX(midy_weight);
DATA_VECTOR(index);
DATA_VECTOR(olandings);
DATA_IVECTOR(iyear);
DATA_IVECTOR(iage);
DATA_IVECTOR(isurvey);
DATA_IVECTOR(isd);
DATA_IVECTOR(is_year);
DATA_IVECTOR(iq);
DATA_VECTOR(fs);
DATA_INTEGER(A);
```

```
DATA_INTEGER(Y);
DATA_INTEGER(Ns);
DATA_ARRAY(crl);
DATA_IVECTOR(isurvey1);
DATA_VECTOR(landings_L);
DATA_VECTOR(landings_U);
DATA_VECTOR(log_landings);
DATA_VECTOR(log_landings_L);
DATA_VECTOR(log_landings_U);
DATA_SCALAR(std_log_landings);
DATA_VECTOR(log_lowerM);
DATA_VECTOR(log_upperM);
DATA_INTEGER(use_pe);

int n = index.size();
Type one = 1.0;
Type zero = 0.0;

//define fixed parameters;
PARAMETER_VECTOR(log_R);
PARAMETER_VECTOR(log_qparm);
PARAMETER_VECTOR(log_cv_index);
PARAMETER(log_std_log_R);
PARAMETER_MATRIX(log_F_mean);
PARAMETER_VECTOR(log_std_log_F);
PARAMETER_ARRAY(log_std_pe);
PARAMETER_ARRAY(log_std_crl);
PARAMETER_VECTOR(logit_ar_index_age);
PARAMETER_VECTOR(logit_ar_logF);
PARAMETER_VECTOR(logit_ar_crl);
PARAMETER_VECTOR(logit_ar_pe);
PARAMETER(logit_ar_logRec);

//define random effects
PARAMETER_ARRAY(log_F_devt);
PARAMETER_ARRAY(log_Nt);
PARAMETER_VECTOR(resid_index);
PARAMETER_MATRIX(resid_crl);

array<Type> log_F_dev = log_F_devt.transpose();
array<Type> log_N = log_Nt.transpose();

//  set bounds on parameters
```

```
Type std_log_R = exp(log_std_log_R);
vector <Type> std_log_F = exp(log_std_log_F);
vector<Type> cv_index = exp(log_cv_index);
//  Type std_pe = exp(log_std_pe);

//  Type std_crl = exp(log_std_crl);
Type ar_logF_age = invlogit(logit_ar_logF(0));
Type ar_logF1_year = invlogit(logit_ar_logF(1));
Type ar_logF_year = invlogit(logit_ar_logF(2));
Type ar_crl_age = invlogit(logit_ar_crl(0));
Type ar_crl_year = invlogit(logit_ar_crl(1));
Type ar_pe_age = invlogit(logit_ar_pe(0));
Type ar_pe_year = invlogit(logit_ar_pe(1));
Type ar_logRec = invlogit(logit_ar_logRec);
vector<Type> ar_index_age = exp(logit_ar_index_age)/
(one + exp(logit_ar_index_age));

//containers
matrix<Type> N(Y,A);
array<Type> pe(Y-1,A-1);
vector<Type> log_Rec_dev(Y);
vector<Type> log_Rec(Y);

matrix<Type> EC(Y,A-4);
matrix<Type> ECW(Y,A-4);

vector<Type> C_tot(Y);
vector<Type> CW_tot(Y);
vector<Type> log_landings_pred(Y);
vector<Type> std_landings_resid(Y);
vector<Type> landings_resid(Y);

vector<Type> Elog_index(n);
vector<Type> Eindex(n);
//vector<Type> resid_index(n);
vector<Type> std_resid_index(n);
vector<Type> log_q_vec = log_qparm(iq);
vector<Type> resid_rec(n);

matrix<Type> p_ya(A-4,Y);
matrix<Type> pya_sum(A-4,Y);
matrix<Type> pi_ya(A-4,Y);
matrix<Type> pred_crl(Y,A-5);
```

```
//matrix<Type> resid_crl(Y,A-5);
matrix<Type> std_resid_crl(Y,A-5);
array<Type> std_crl(Y,A-5);
array<Type> std_pe(Y-1,A-1);

//SD report objects
matrix<Type> B_matrix(Y,A);
matrix<Type> SSB_matrix(Y,A-4);
vector<Type> biomass(Y);
vector<Type> log_biomass(Y);
vector<Type> ssb(Y);
vector<Type> log_ssb(Y);
vector<Type> aveF_914(Y);
vector<Type> log_aveF_914(Y);


//initialize the negative log likelihood
Type nll = zero;
using namespace density;

//to model F (no fishing up to age 5)
matrix<Type> F(Y,A-4);
matrix<Type> Z(Y,A);
matrix<Type> log_F = log(F.array());

int i,j;
for(i = 0;i < Y;++i){
for(j = 0;j < A-5;++j){std_crl(i,j) = exp(log_std_crl(i,j));}
}
for(i = 0;i < Y-1;++i){
for(j = 0;j < A-1;++j){std_pe(i,j) = exp(log_std_pe(i,j));}
}
for(i = 0;i < Y;++i){
for(j = 0;j < A-4;++j){
log_F(i,j) = std_log_F(j)*log_F_dev(i,j) + log_F_mean(i,j);
F(i,j) = exp(log_F(i,j));
}
}

for(int i = 0;i < Y;++i){
for(int j = 0;j < A;++j){
if(j<4){Z(i,j) = M(i,j);}
if(j>=4){Z(i,j)= F(i,j-4)+M(i,j);}
```

```
}}

//compute process errors
for(int i = 1;i < Y;++i){
for(int j = 1;j < A-1;++j){
pe(i-1,j-1) = log_N(i,j) - log_N(i-1,j-1) + Z(i-1,j-1);
}
int j=A-1;
pe(i-1,j-1) = log_N(i,j) - log(exp(log_N(i-1,j-1)-Z(i-1,j-1))+
exp(log_N(i-1,j)-Z(i-1,j)));
}

//calculate recruitment deviations;
for(int i = 0;i < Y;++i){
if(i<=32){log_Rec_dev(i)= log_N(i,0) - log_R(0);}
if(i>32){log_Rec_dev(i)= log_N(i,0) - log_R(1);}
log_Rec(i) = log_N(i,0);
}

for(int i = 0;i < Y;++i){
for(int j = 0;j < A;++j){
N(i,j)=exp(log_N(i,j));}}


//Baranov catch equation predictions and residuals
for(int i = 0;i < Y;++i){
C_tot(i) = zero;
CW_tot(i) = zero;
for(int j = 0;j < A-4;++j){
EC(i,j) = N(i,j+4)*((one - exp(-one*Z(i,j+4)))*F(i,j)/Z(i,j+4));
ECW(i,j) = EC(i,j)*midy_weight(i,j);
C_tot(i) += EC(i,j);
CW_tot(i) += ECW(i,j);
} }

log_landings_pred = log(CW_tot);
landings_resid = log_landings - log_landings_pred;
std_landings_resid = landings_resid/std_log_landings;

//age composition catch
for(int i = 0;i < Y;++i){
for(int j = 0;j < A-4;++j){
p_ya(j,i) = EC(i,j)/C_tot(i);
```

```
}}

Type total;
for(int i = 0;i <Y;++i){
total=zero;
pya_sum(0,i) = one;
for(int j = 1;j < A-4;++j){
total+=p_ya(j-1,i);
pya_sum(j,i) = (one-total);
}}

for(int i = 0;i <Y;++i){
for(int j = 0;j < A-4;++j){
pi_ya(j,i) = p_ya(j,i)/pya_sum(j,i);}}


for(int i = 0;i<Y;i++){
for(int j = 0;j<A-5;j++){
pred_crl(i,j)= log(pi_ya(j,i)/(one - pi_ya(j,i)));
//resid_crl(i,j) = crl(i,j) - pred_crl(i,j);
std_resid_crl(i,j) = resid_crl(i,j)/std_crl(i,j);}}

//Survey index predictions, and residuals;

vector<Type> std_index_vec(n);
matrix<Type> mresid_index(Ns,A);
matrix<Type> msd_index(Ns,A);

int ia,iy,iy1,is;
for(i = 0;i < n;++i){
ia = iage(i);
iy = iyear(i);
is = isd(i);
iy1 = is_year(i);

Elog_index(i) = log_q_vec(i) + log_N(iy,ia) - fs(i)*Z(iy,ia);
Eindex(i) = exp(Elog_index(i));
std_index_vec(i) = cv_index(is)*Eindex(i);
//resid_index(i) = index(i) - Eindex(i);
std_resid_index(i) = resid_index(i)/std_index_vec(i);

mresid_index(iy1,ia) = resid_index(i);
msd_index(iy1,ia) = std_index_vec(i);
```

```
}

//NEGATIVE LOGLIKELIHOODS
//Index OBSERVATION MODEL

vector<Type> del(A);
vector<Type> sd_del(A);

for(i = 0;i < Ns;++i){
iy = isurvey1(i);
del = vector<Type>(mresid_index.row(i));
sd_del = vector<Type>(msd_index.row(i));
nll += VECSCALE(AR1(ar_index_age(iy)),sd_del)(del);
}

array<Type> temp(Y,A-5);
temp = resid_crl.array();
nll += VECSCALE(SEPARABLE(AR1(ar_crl_age),AR1(ar_crl_year)),std_crl)(temp);

ADREPORT(resid_index);
ADREPORT(resid_crl);

return nll;
}
```